TUGAS MINGGU KE-13

STATISTIKA DESKRIPTIF

NAMA : MUKHAMAD IKHSANUDIN

NIM : 082011633086

PROGRAM STUDI S1 SISTEM INFORMASI

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS AIRLANGGA

2021

Tugas pertemuan 26 → dikumpulkan hari ini, tgl. 11-06-2021 jam 23.59 → di upload ke Aula dan kirim ke email eto-w@fst.unair .ac.id dengan subject : Clustering

Code dan outputnya jadikan satu di notebook R-nya

1. Carilah **3 dataset** yang sesuai untuk Clustering kemudian gunakan metode :
   a. 5 metode agglomerative (centroid, single-linkage, complete-linkage, average-linkage dan ward)
   b. Tentukan nilai k (banyaknya klaster) berdasarkan dendogramnya
   c. Tentukan nilai k (banyaknya klaster) berdasarkan nilai BIC atau yang lain
   d. Buatlah table rekapitulasi

No	Metode	Nomer klaster	Anggota klaster
1	Centroid	1	
		2	
		k	
2	single-linkage		

2. Gunakan no 1 untuk metode-metode berikut :
   a. K-means dan buatlah table rekapitulasi seperti no 1d.
   b. K-medoids dan buatlah table rekapitulasi seperti no 1d.
   c. K-medians dan buatlah table rekapitulasi seperti no 1d.

==========

Code ditaruh diantara tanda berikut :

```{R}

Syntax di sini

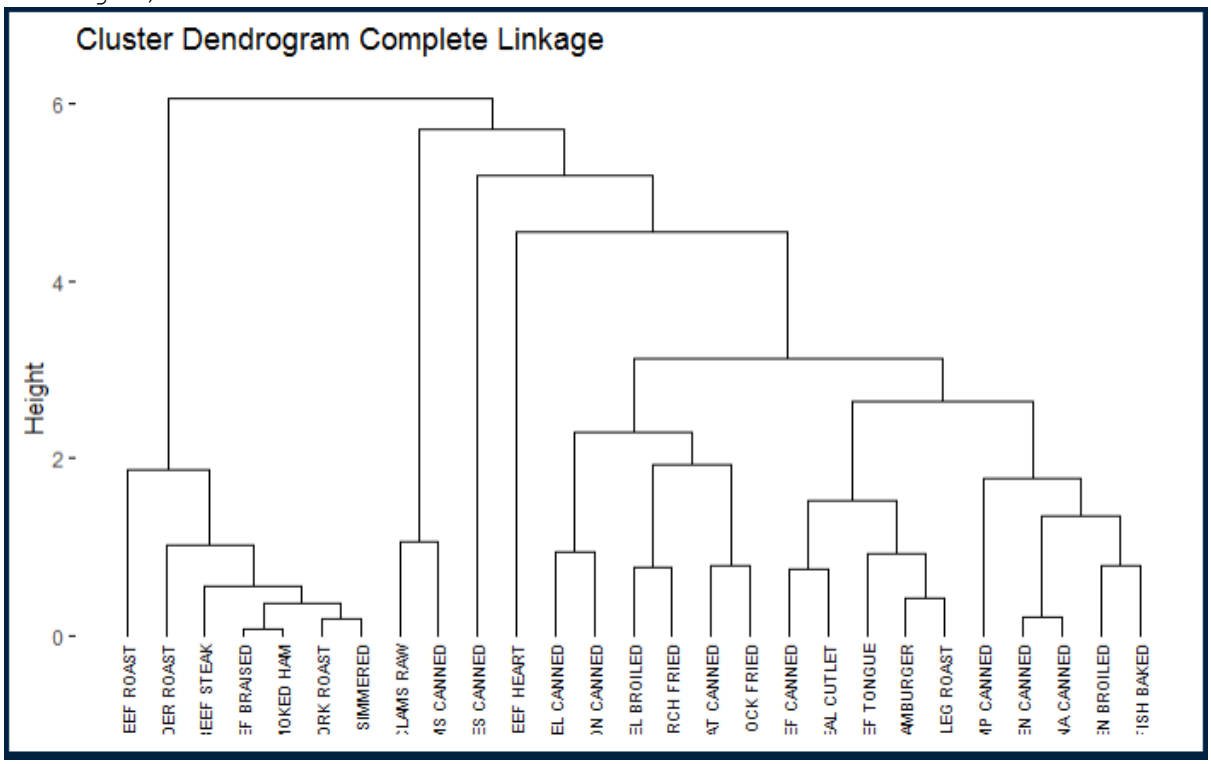
```

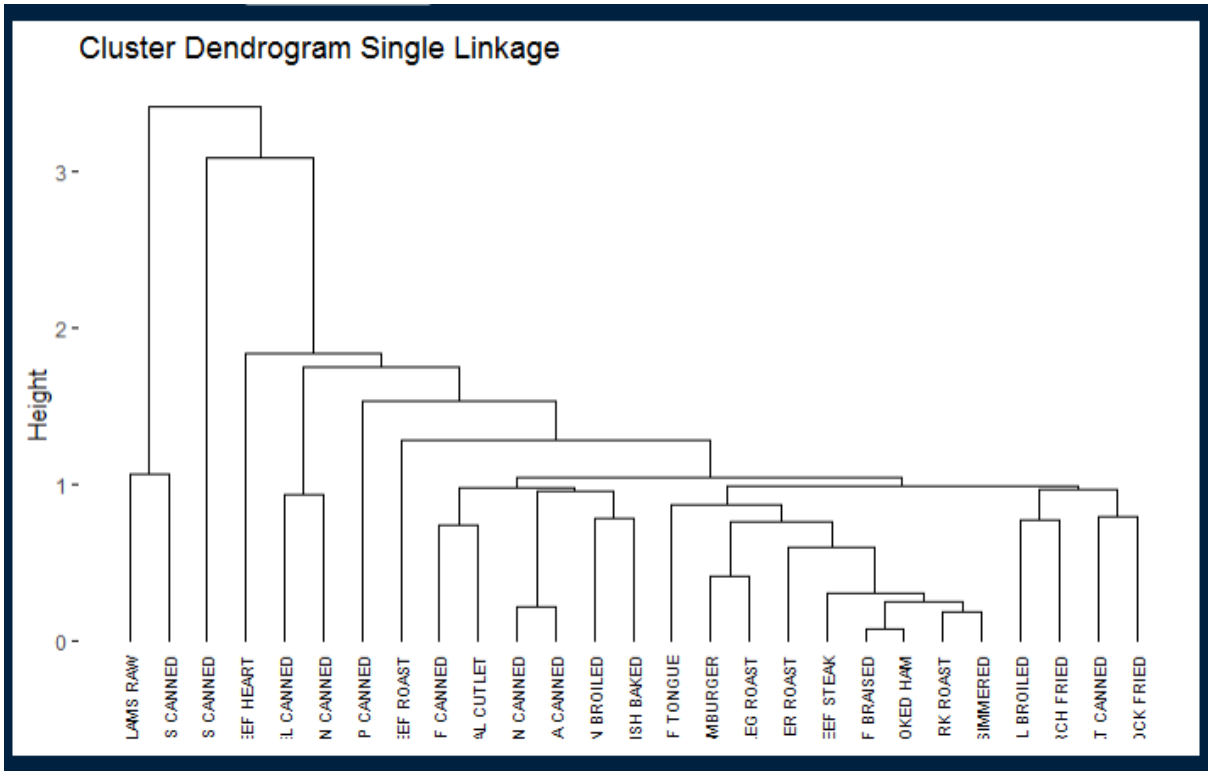
Carilah **3 dataset** yang sesuai untuk Clustering kemudian gunakan metode :

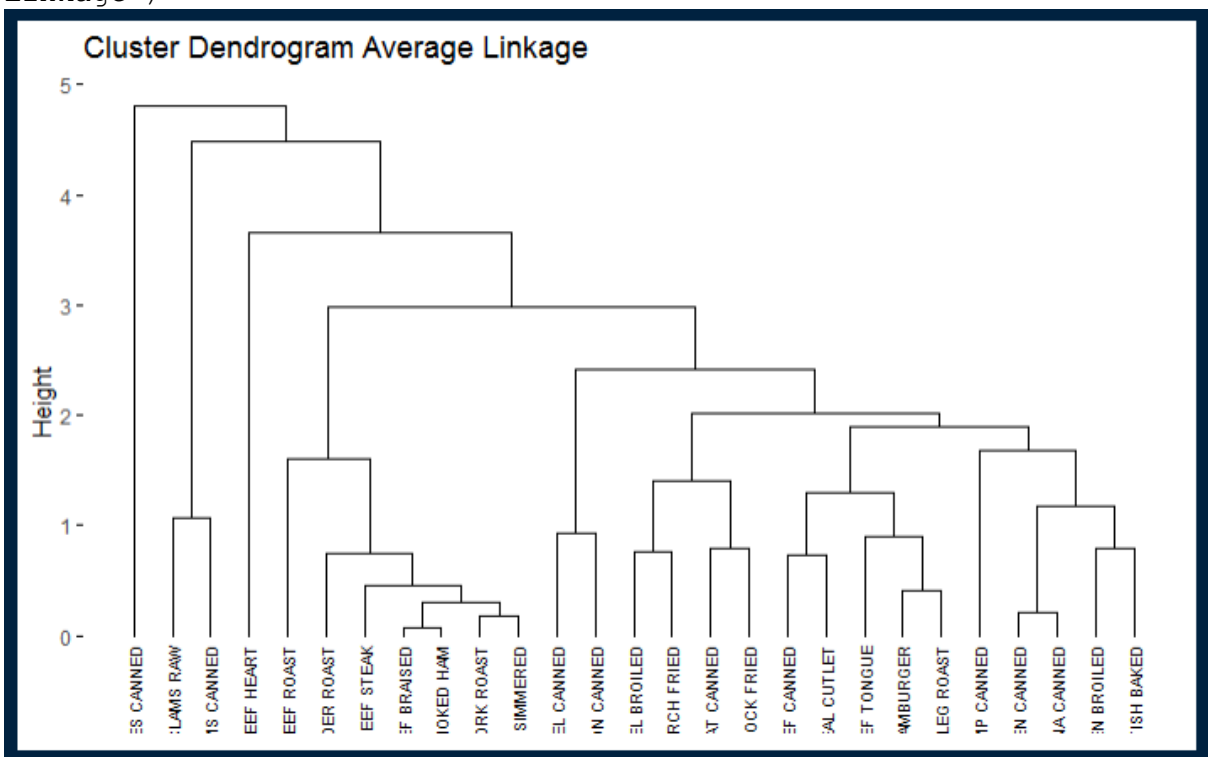
a. **5 metode agglomerative (centroid, single-linkage, complete-linkage, average-linkage dan ward)**

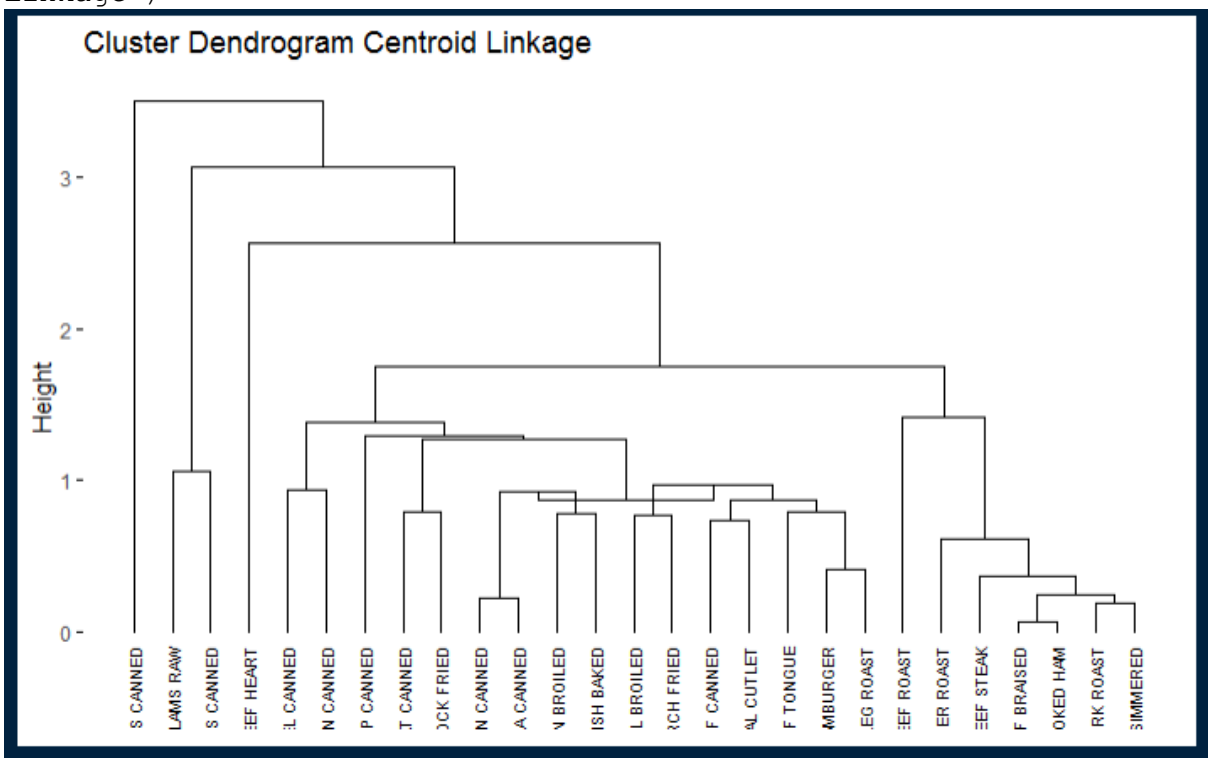
```
library(tidyverse)
library(cluster.datasets)
library(factoextra)

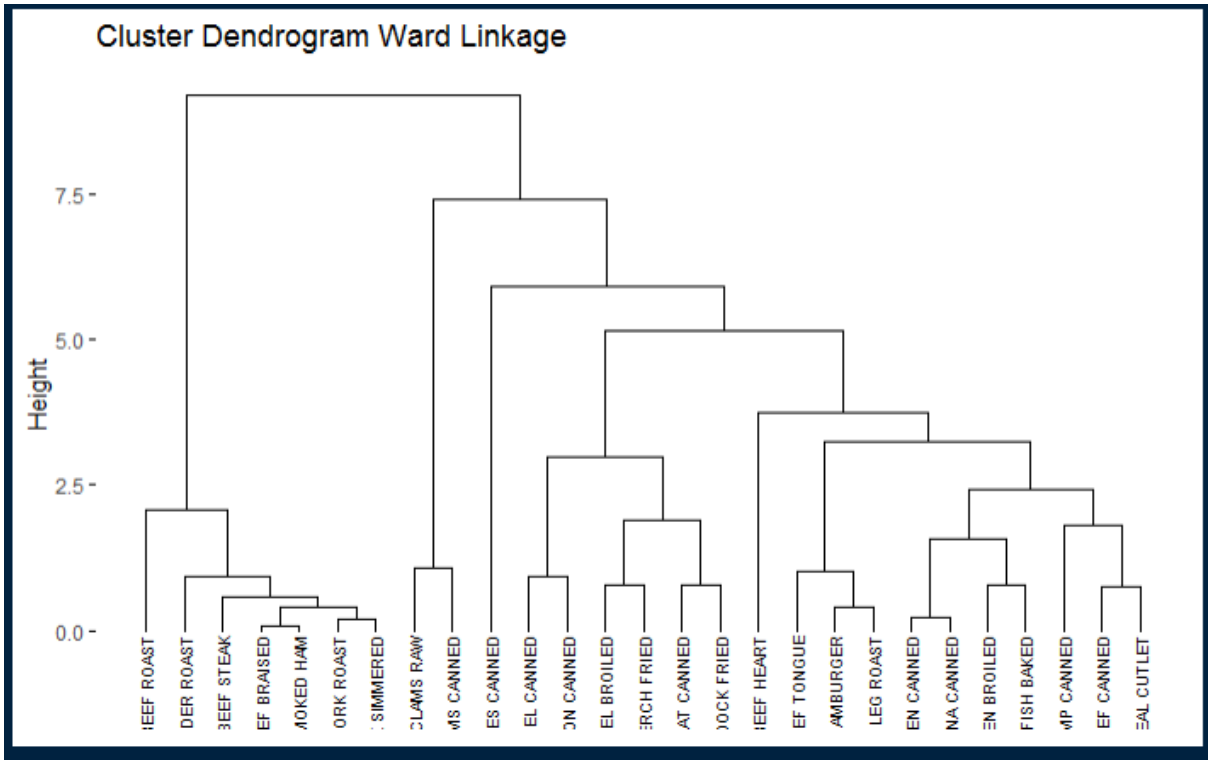
```{r}
# Dataset 1

data("nutrient")
DataClus1 <- nutrient
summary(DataClus1)
DataClus1_Fix <- scale(DataClus1)
DataClus1_Fix

Clus1Dist <- dist(x = DataClus1_Fix, method = "euclidean")

# Complete Linkage
Clus1_Com <- hclust(d = Clus1Dist, method = "complete")
Clus1_Com
fviz_dend(Clus1_Com, cex = 0.5, main = "Cluster Dendrogram Complete
Linkage")
```




```
# Single Linkage
Clus1_Sin <- hclust(d = Clus1Dist, method = "single")
Clus1_Sin
fviz_dend(Clus1_Sin, cex = 0.5, main = "Cluster Dendrogram Single
Linkage")
```
```




```
# Average Linkage
Clus1_Ave <- hclust(d = Clus1Dist, method = "average")
Clus1_Ave

fviz_dend(Clus1_Ave, cex = 0.5, main = "Cluster Dendrogram Average
Linkage")
```



```
# Centroid Linkage
Clus1_Cen <- hclust(d = Clus1Dist, method = "centroid")
Clus1_Cen

fviz_dend(Clus1_Cen, cex = 0.5, main = "Cluster Dendrogram Centroid
Linkage")
```




```
# Ward Linkage
Clus1_War <- hclust(d = Clus1Dist, method = "ward.D2")
Clus1_War

fviz_dend(Clus1_War, cex = 0.5, main = "Cluster Dendrogram Ward
Linkage")
```




```

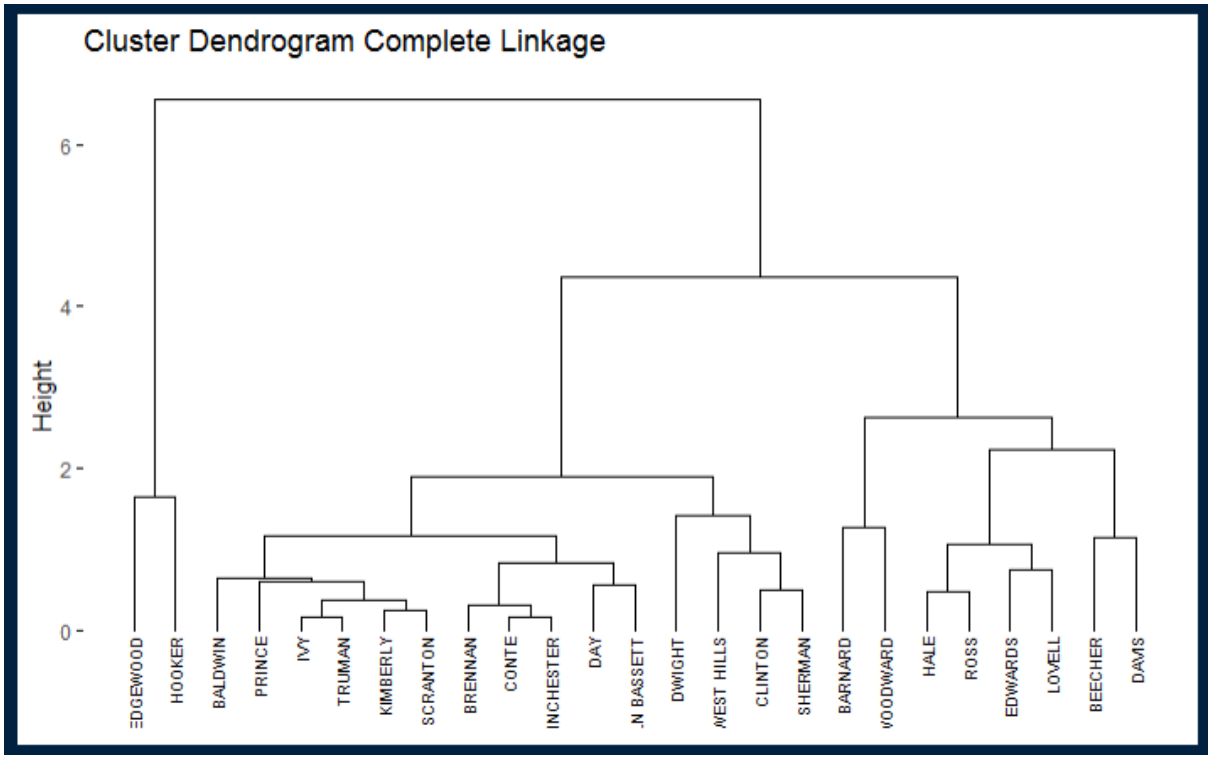
```{R}
#Dataset 2

data("achieve")
DataClus2 <- achieve
summary(DataClus2)
DataClus2_Fix <- scale(DataClus2)
DataClus2_Fix

Clus2Dist <- dist(x = DataClus2_Fix, method = "euclidean")

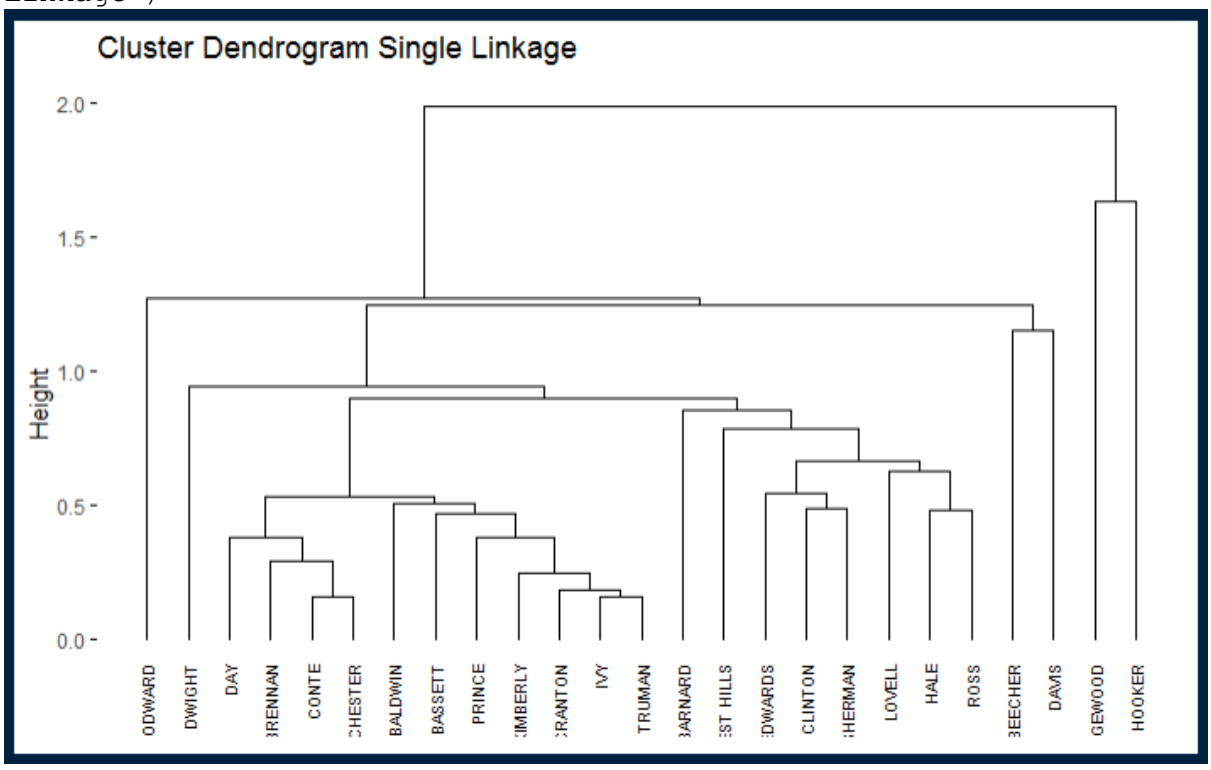
# Complete Linkage
Clus2_Com <- hclust(d = Clus2Dist, method = "complete")
Clus2_Com

fviz_dend(Clus2_Com, cex = 0.5, main = "Cluster Dendrogram Complete
Linkage")
```
```



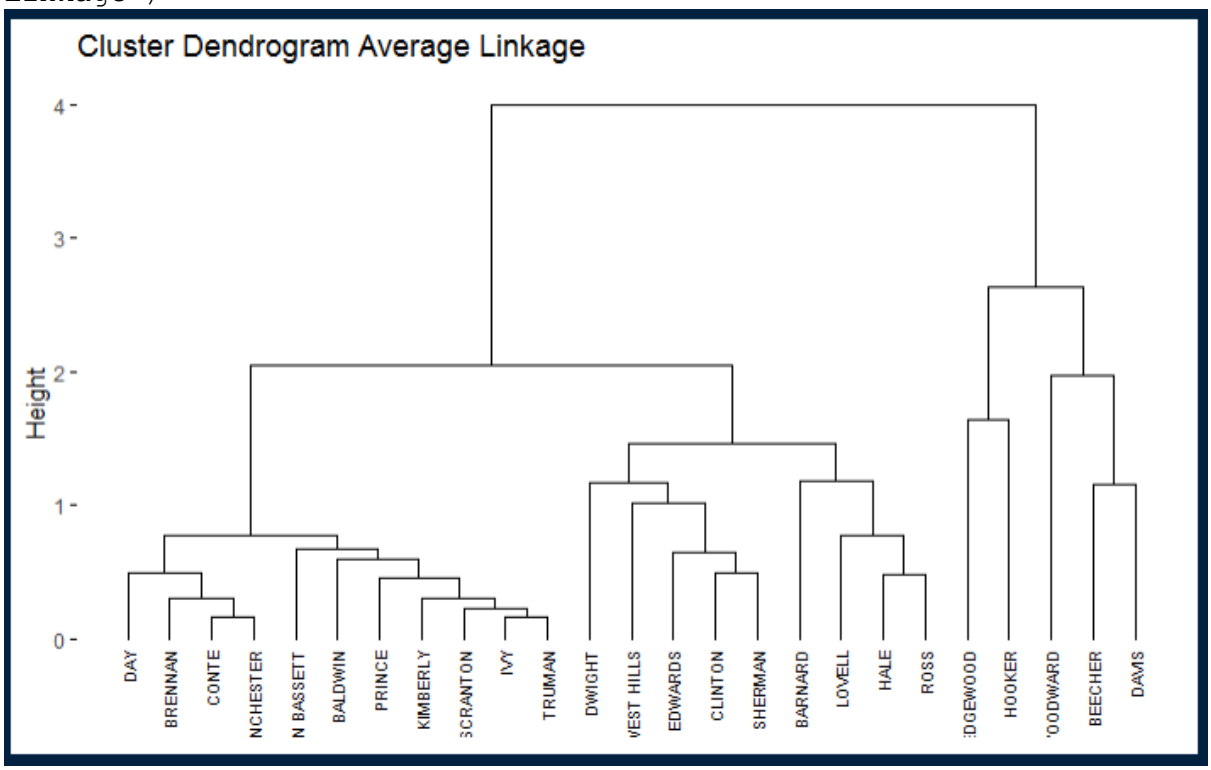

```
# Single Linkage
Clus2_Sin <- hclust(d = Clus2Dist, method = "single")
Clus2_Sin

fviz_dend(Clus2_Sin, cex = 0.5, main = "Cluster Dendrogram Single
Linkage")
```



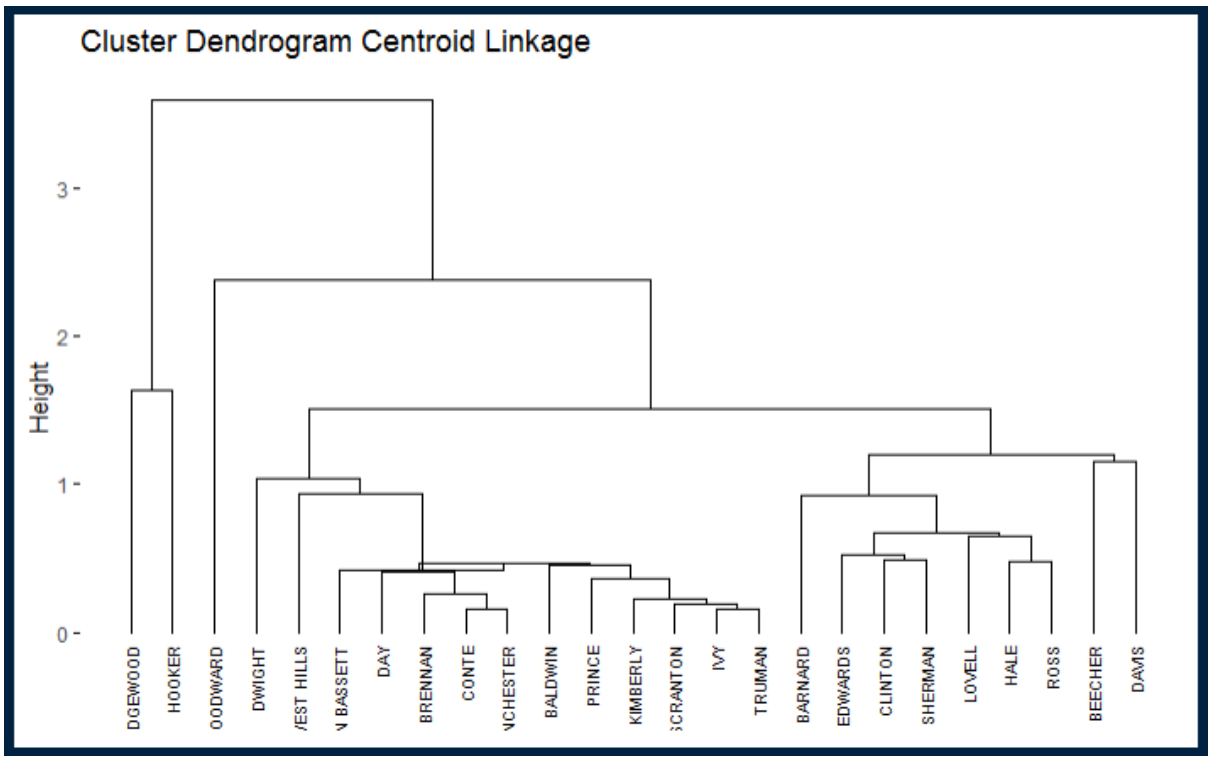
```
# Average Linkage
Clus2_Ave <- hclust(d = Clus2Dist, method = "average")
Clus2_Ave

fviz_dend(Clus2_Ave, cex = 0.5, main = "Cluster Dendrogram Average
Linkage")
```



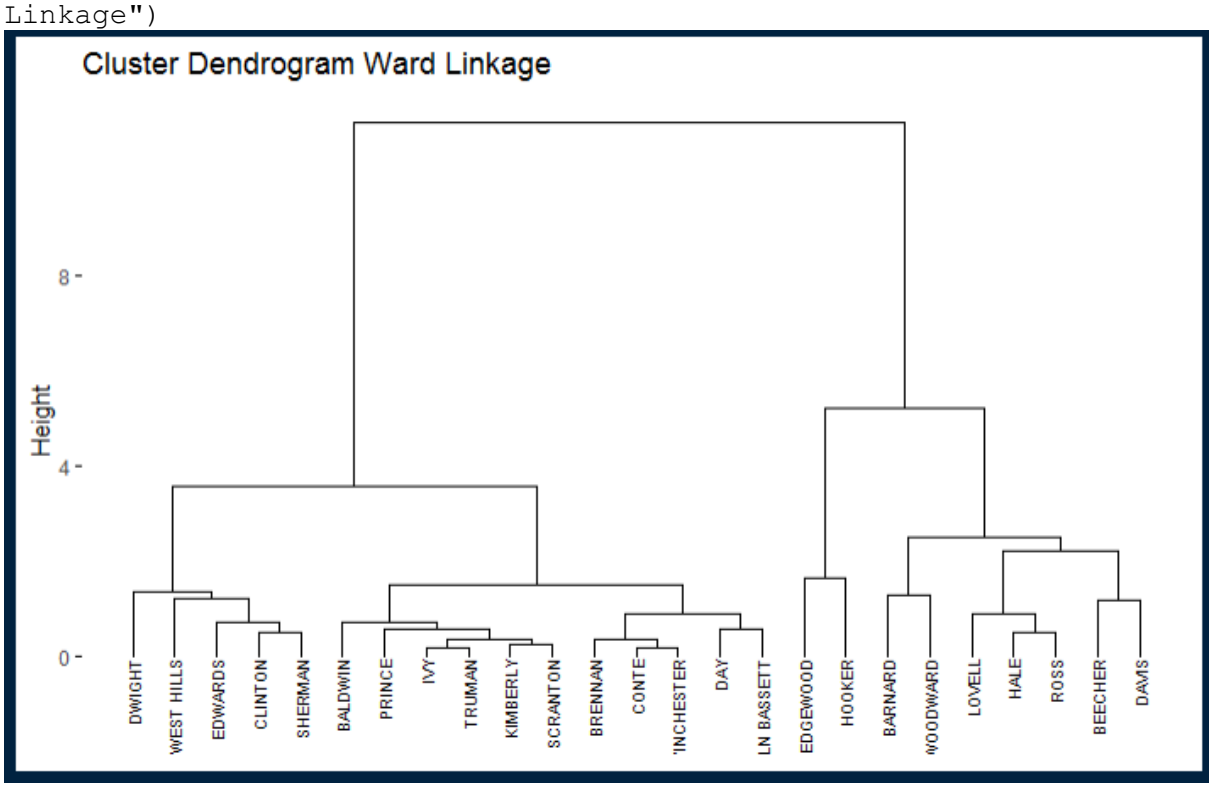

```
# Centroid Linkage
Clus2_Cen <- hclust(d = Clus2Dist, method = "centroid")
Clus2_Cen

fviz_dend(Clus2_Cen, cex = 0.5, main = "Cluster Dendrogram Centroid
Linkage")
```



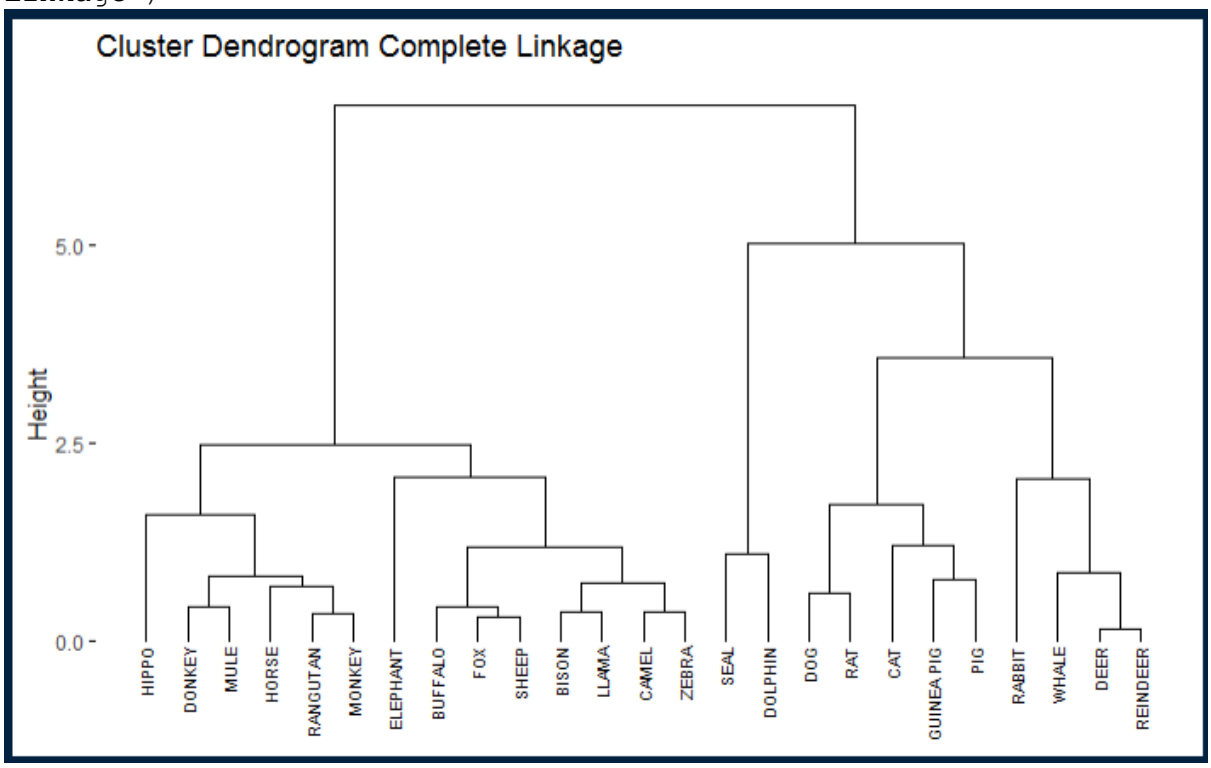

```
# Ward Linkage
Clus2_War <- hclust(d = Clus2Dist, method = "ward.D2")
Clus2_War

fviz_dend(Clus2_War, cex = 0.5, main = "Cluster Dendrogram Ward
Linkage")
```

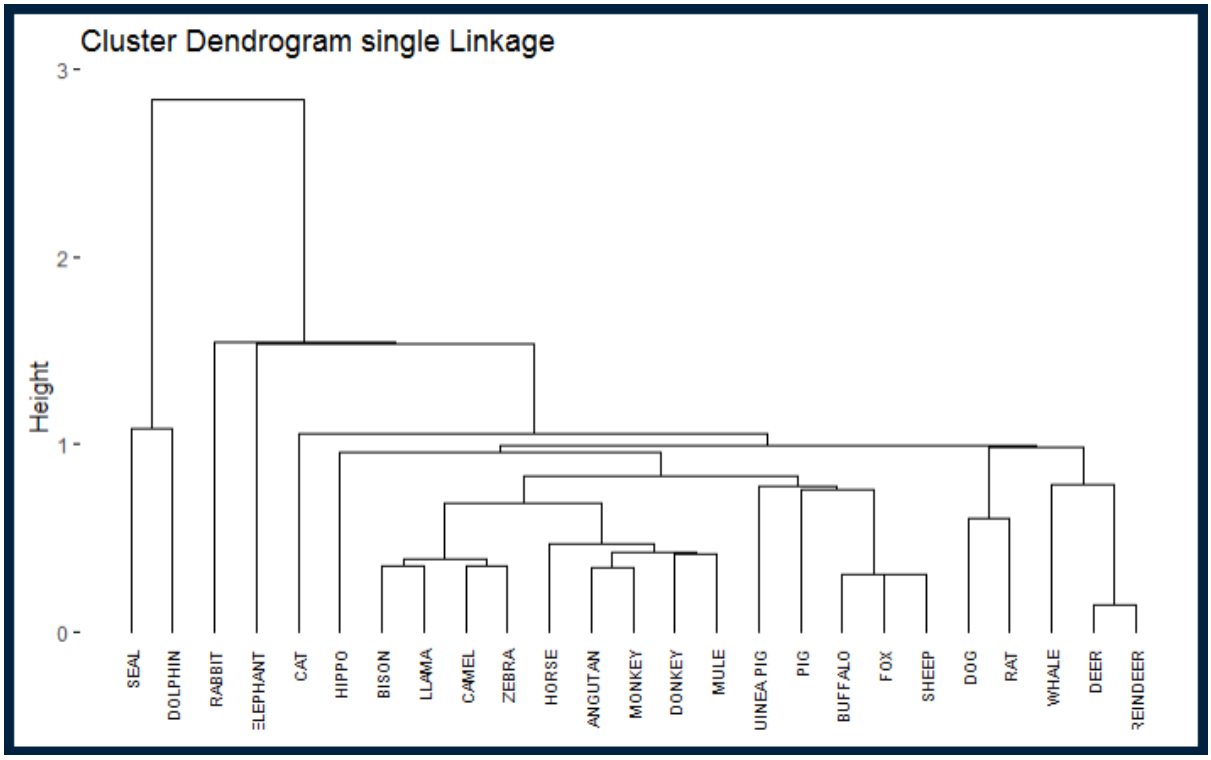



```
```

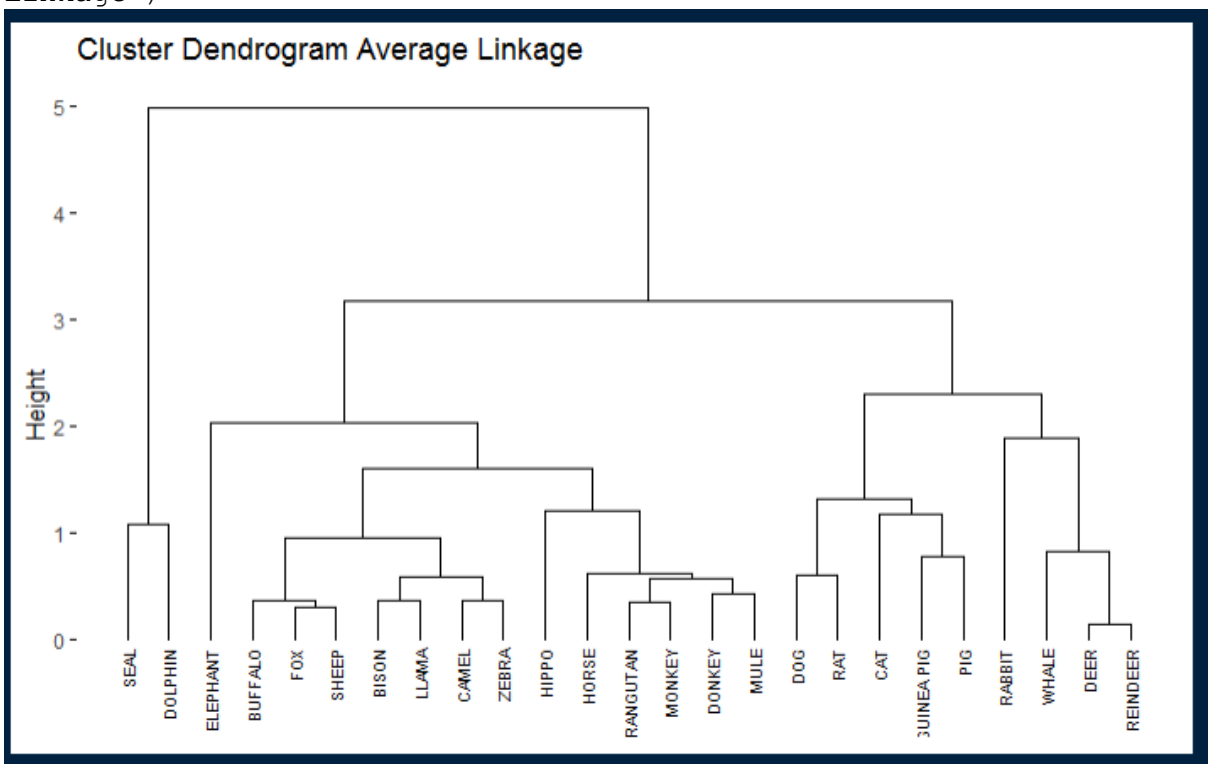
```{R}
#Dataset 3

data("milk")
DataClus3 <- milk
summary(DataClus3)
DataClus3_Fix <- scale(DataClus3)
DataClus3_Fix

Clus3Dist <- dist(x = DataClus3_Fix, method = "euclidean")

# Complete Linkage
Clus3_Com <- hclust(d = Clus3Dist, method = "complete")
Clus3_Com

fviz_dend(Clus3_Com, cex = 0.5, main = "Cluster Dendrogram Complete
Linkage")
```



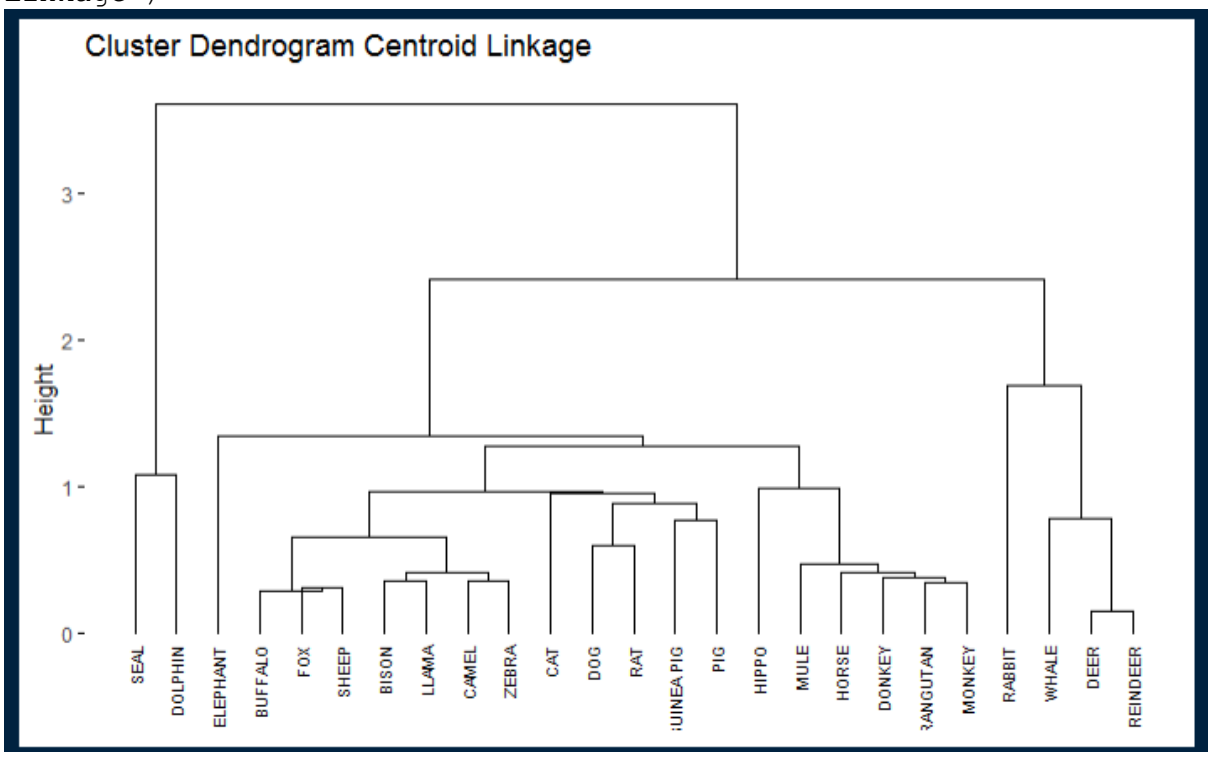

```{R}
# Single Linkage
Clus3_Sin <- hclust(d = Clus3Dist, method = "single")
Clus3_Sin

fviz_dend(Clus3_Sin, cex = 0.5, main = "Cluster Dendrogram single
Linkage")
```



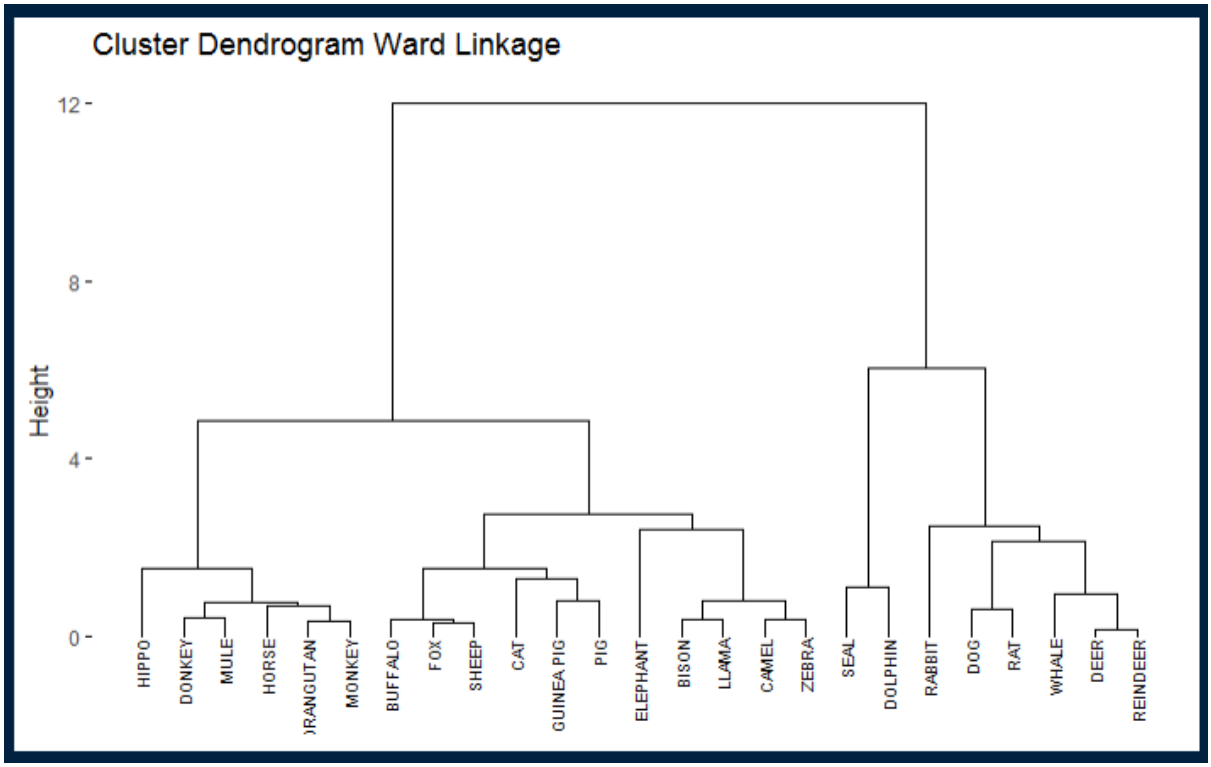

```
# Average Linkage
Clus3_Ave <- hclust(d = Clus3Dist, method = "average")
Clus3_Ave

fviz_dend(Clus3_Ave, cex = 0.5, main = "Cluster Dendrogram Average
Linkage")
```



```
# Centroid Linkage
Clus3_Cen <- hclust(d = Clus3Dist, method = "centroid")
Clus3_Cen

fviz_dend(Clus3_Cen, cex = 0.5, main = "Cluster Dendrogram Centroid
Linkage")
```

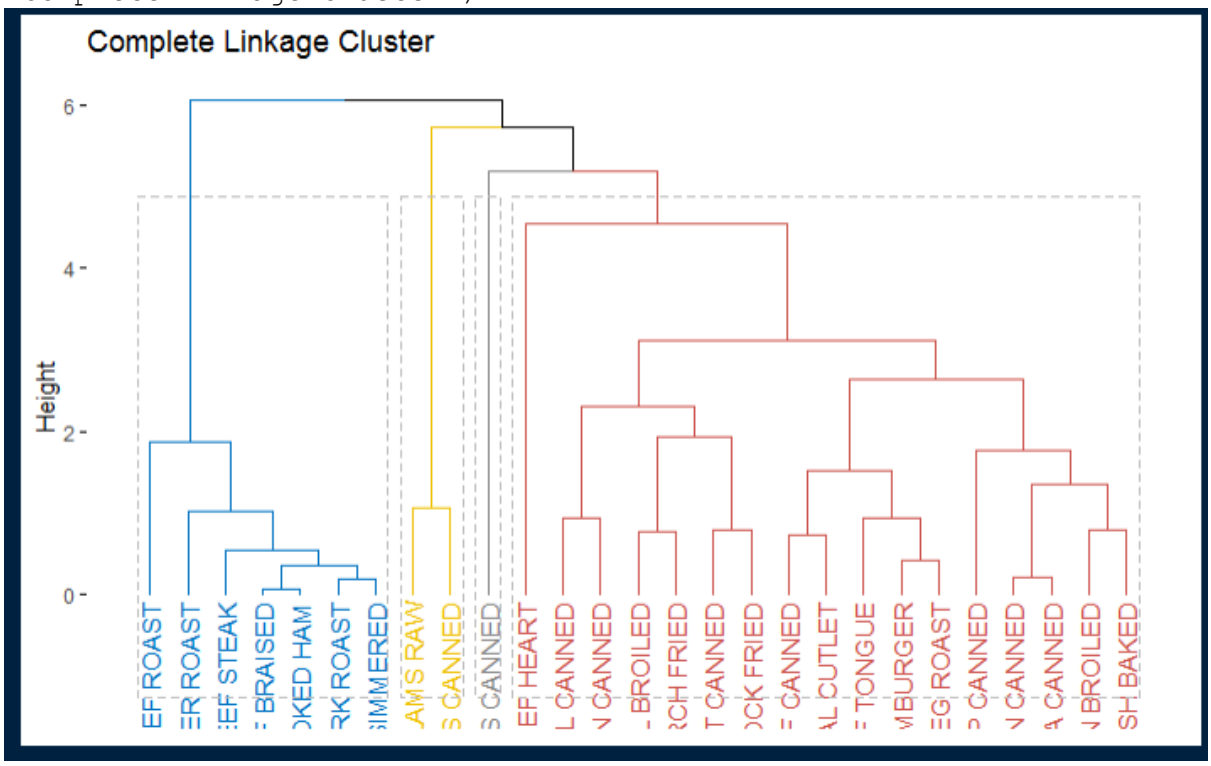



```
# Ward Linkage
Clus3_War <- hclust(d = Clus3Dist, method = "ward.D2")
Clus3_War

fviz_dend(Clus3_War, cex = 0.5, main = "Cluster Dendrogram Ward
Linkage")
```

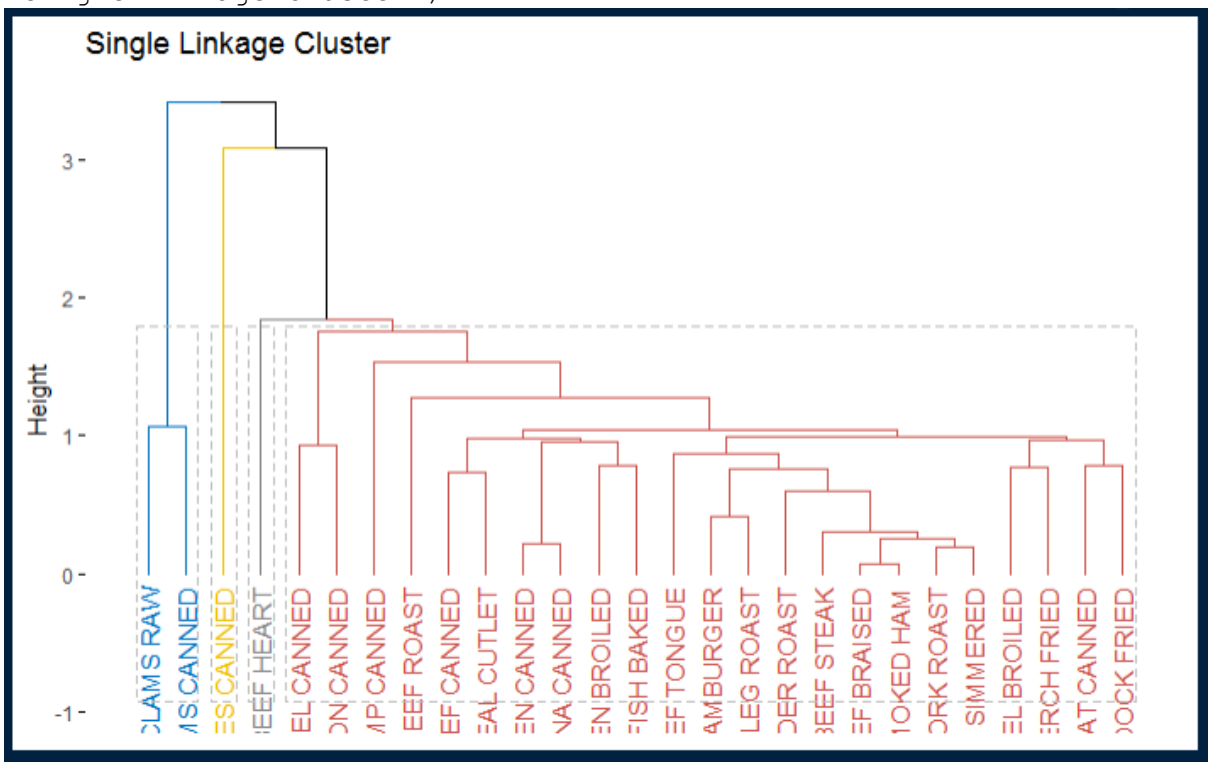

Cluster Dendrogram Ward Linkage
Height
12
8
4
0
HIPPO
DONKEY
MULE
HORSE
ORANGUTAN
MONKEY
BUFFALO
FOX
SHEEP
CAT
GUINEA PIG
PIG
ELEPHANT
BISON
LLAMA
CAMEL
ZEBRA
SEAL
DOLPHIN
RABBIT
DOG
RAT
WHALE
DEER
REINDEER

b. Tentukan nilai k (banyaknya klaster) berdasarkan dendogramnya

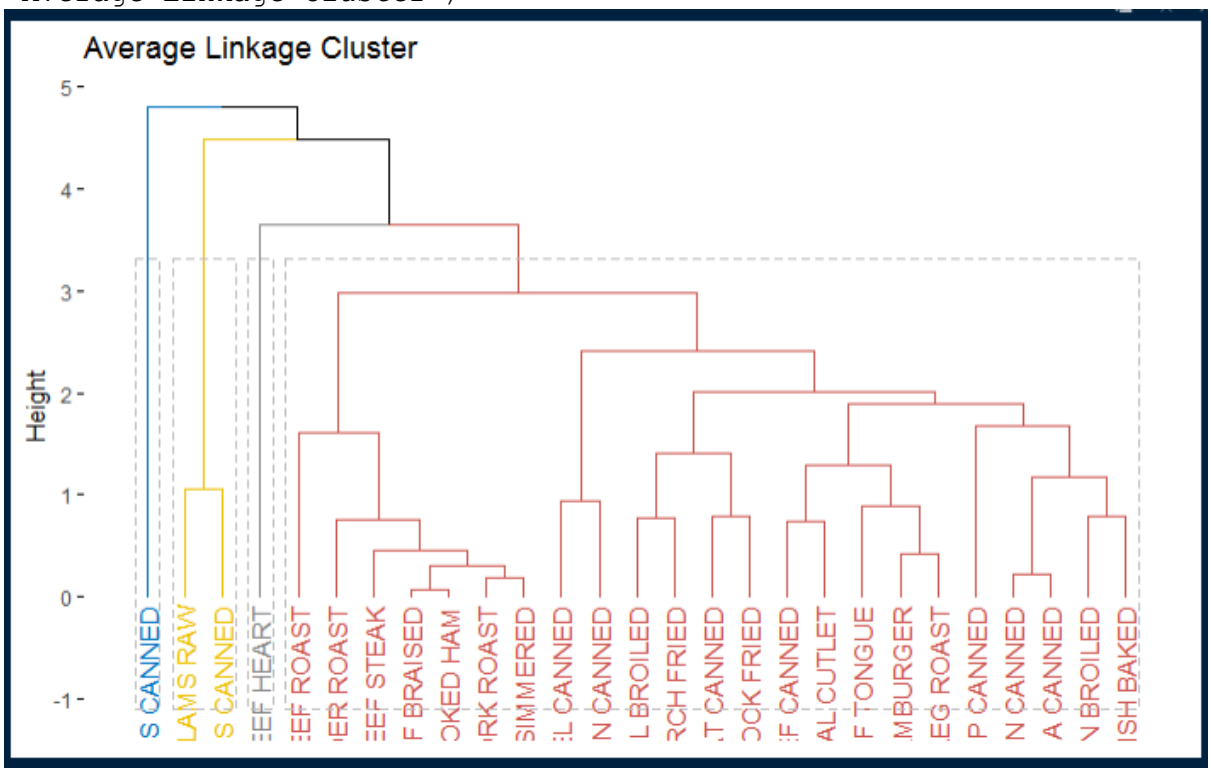
```{R}
#Dataset 1
fviz_dend(Clus1_Com, k = 4, k_colors = "jco", rect = T, main =
"Complete Linkage Cluster")
```



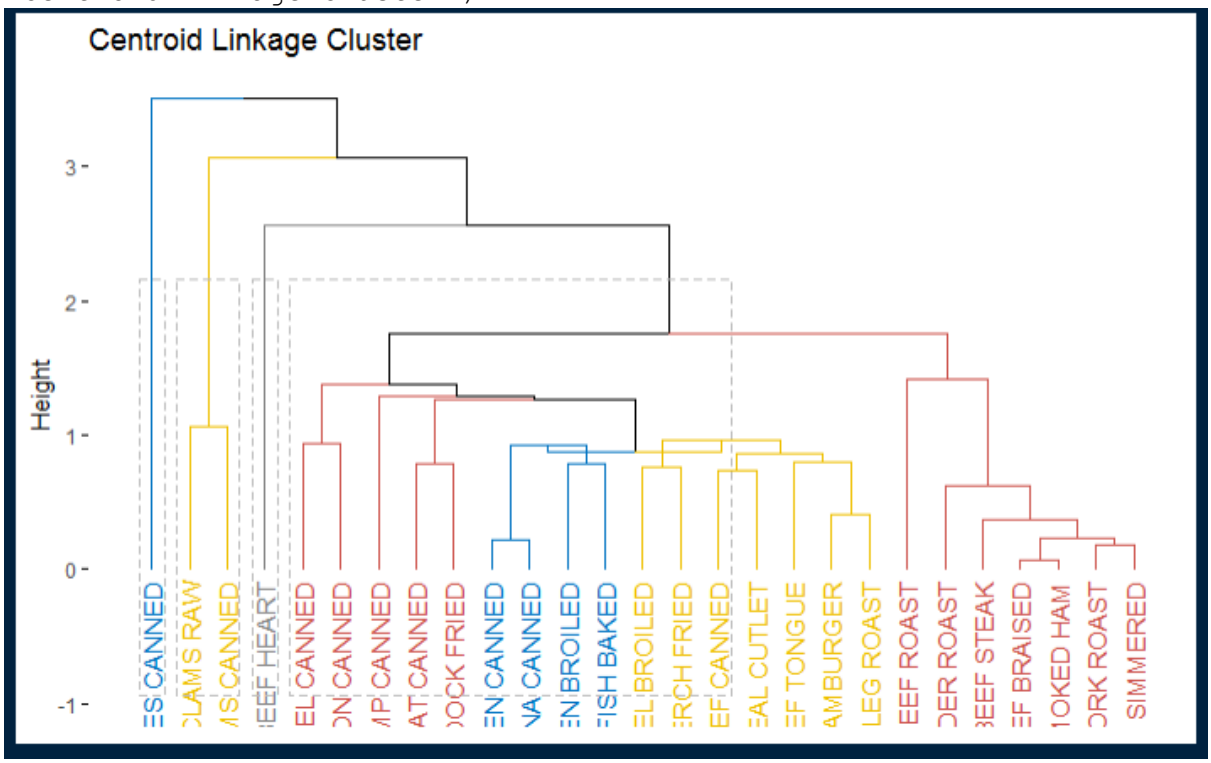

```
fviz_dend(Clus1_Sin, k = 4, k_colors = "jco", rect = T, main =
"Single Linkage Cluster")
```



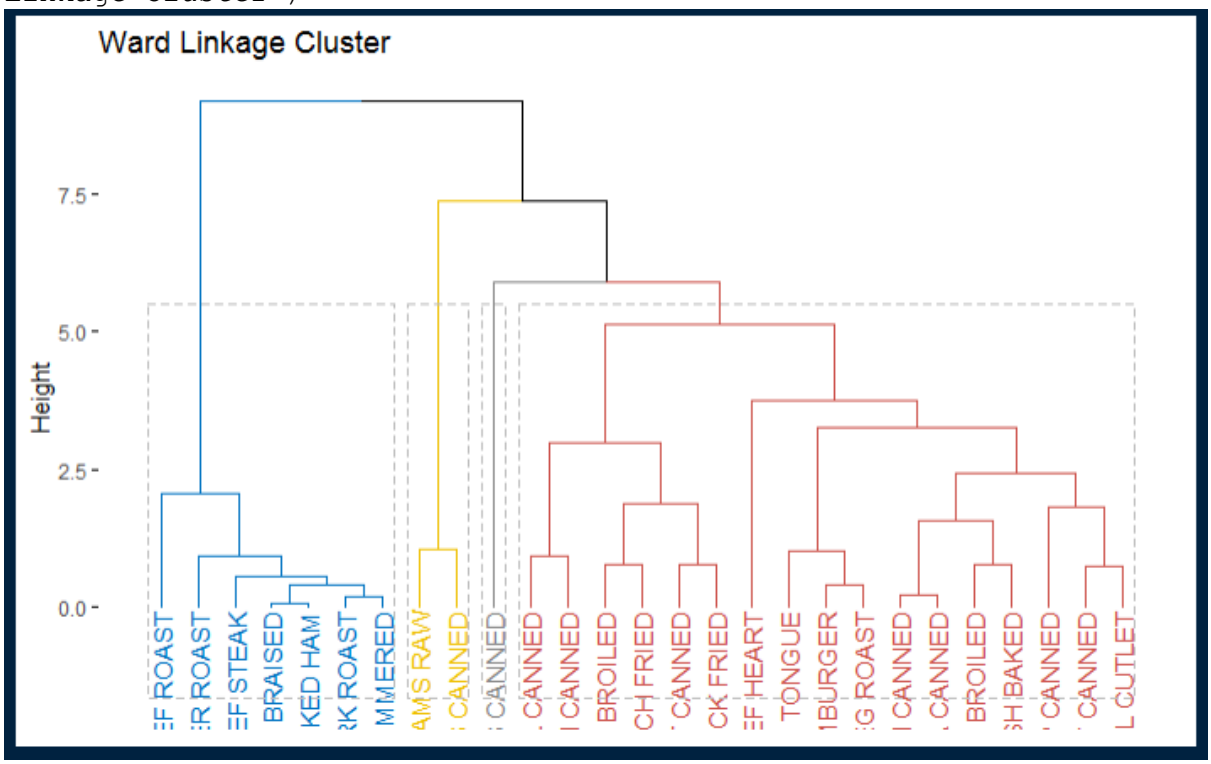
```
fviz_dend(Clus1_Ave, k = 4, k_colors = "jco", rect = T, main =
"Average Linkage Cluster")
```




```
fviz_dend(Clus1_Cen, k = 4, k_colors = "jco", rect = T, main =
"Centroid Linkage Cluster")
```

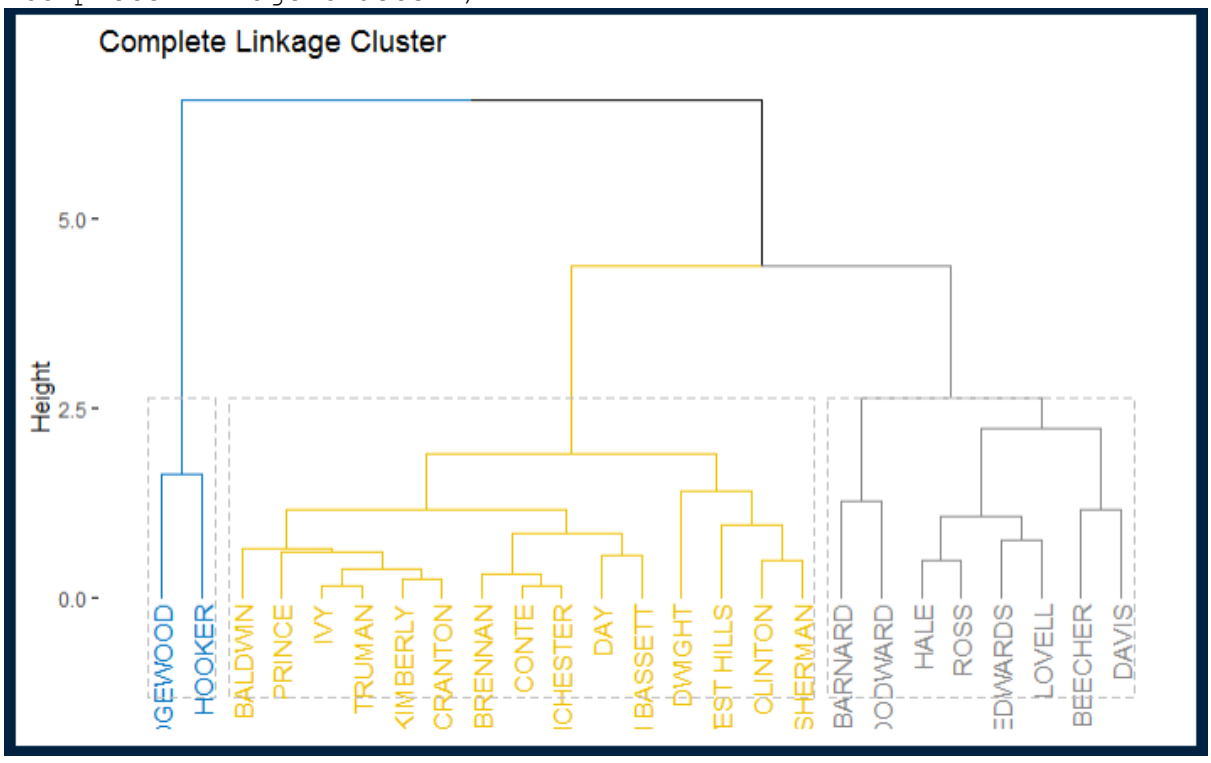


```
fviz_dend(Clus1_War, k = 4, k_colors = "jco", rect = T, main = "Ward
Linkage Cluster")
```

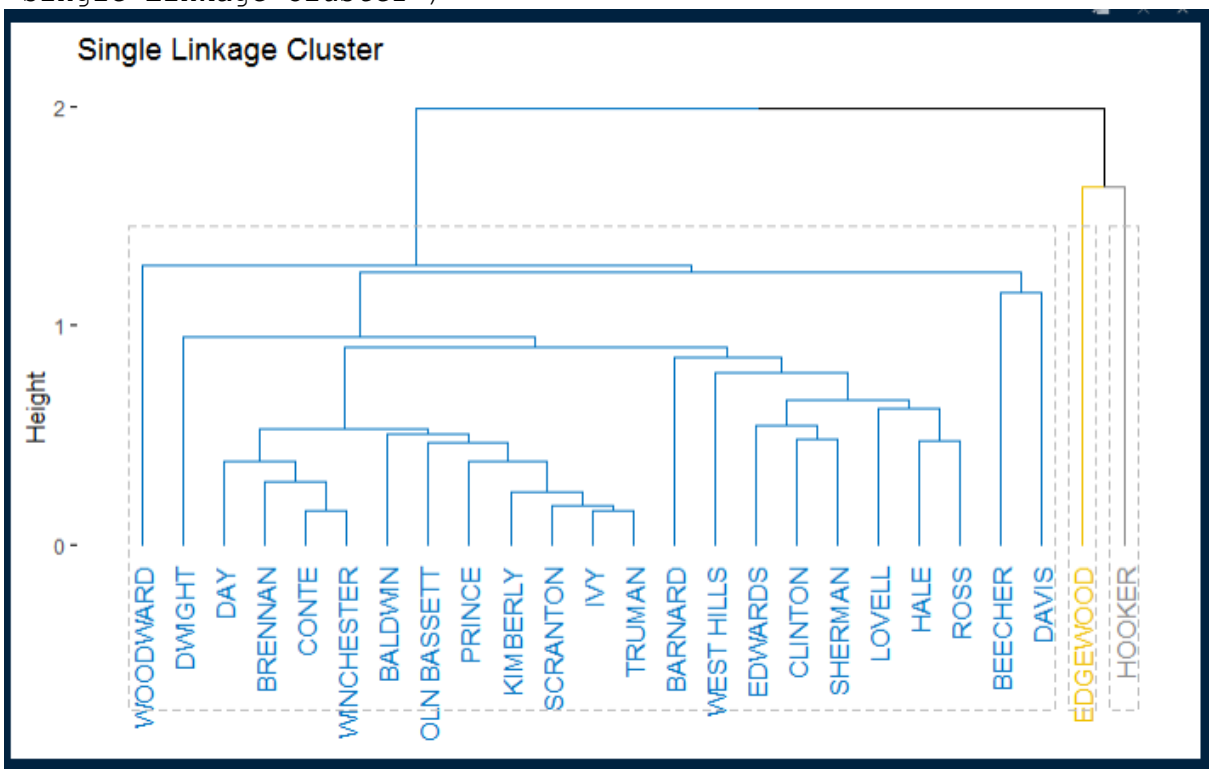



```
```

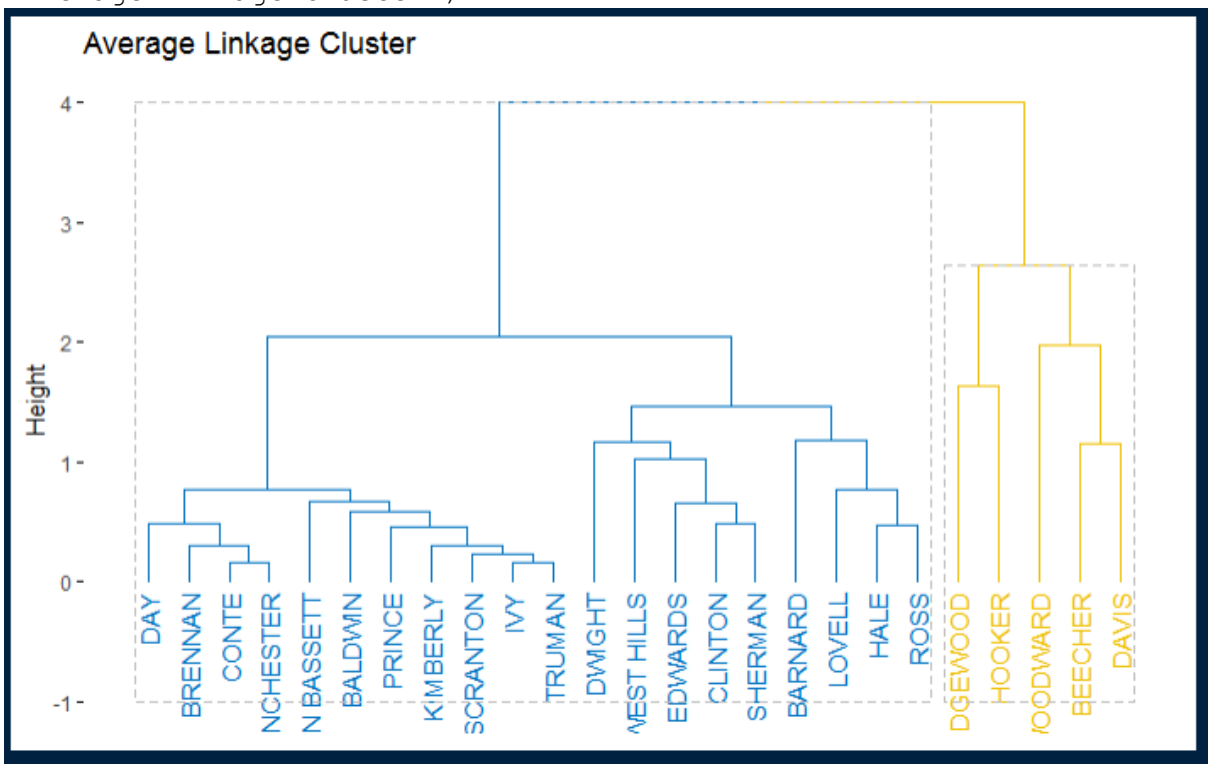
```{R}
#Dataset 2
fviz_dend(Clus2_Com, k = 3, k_colors = "jco", rect = T, main =
"Complete Linkage Cluster")
```



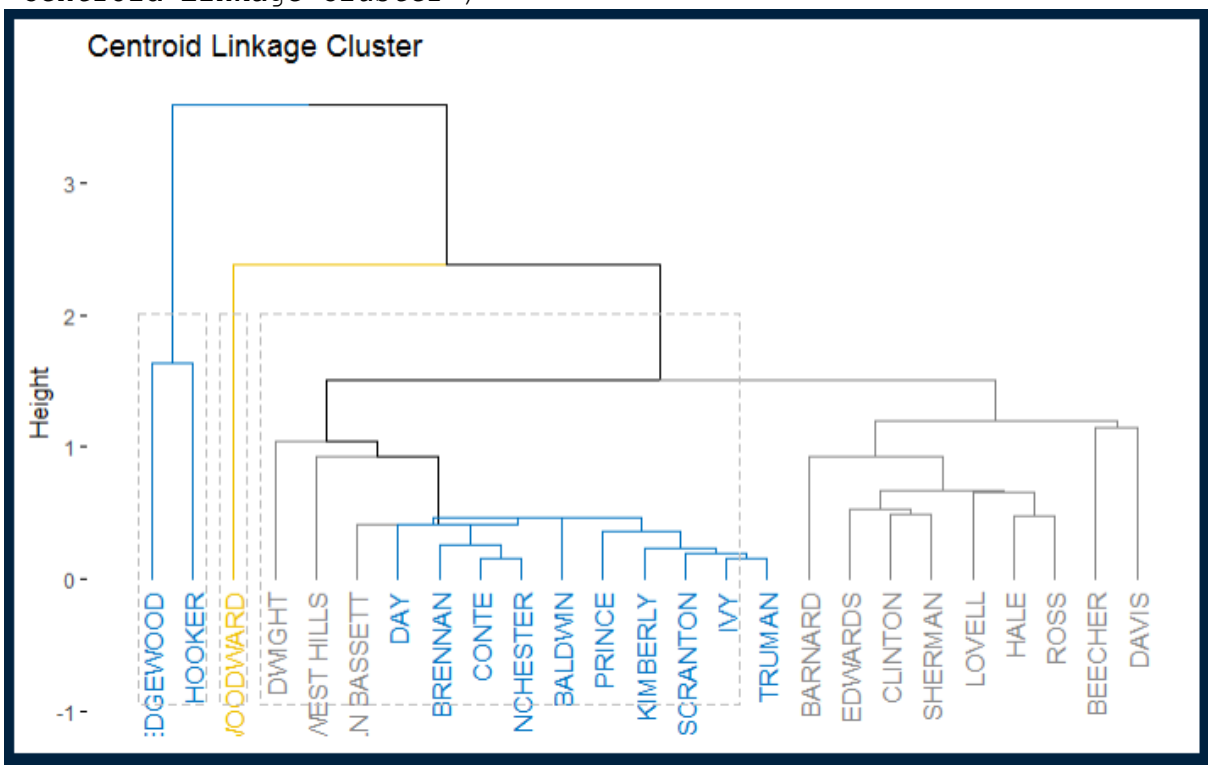
```
fviz_dend(Clus2_Sin, k = 3, k_colors = "jco", rect = T, main =
"Single Linkage Cluster")
```



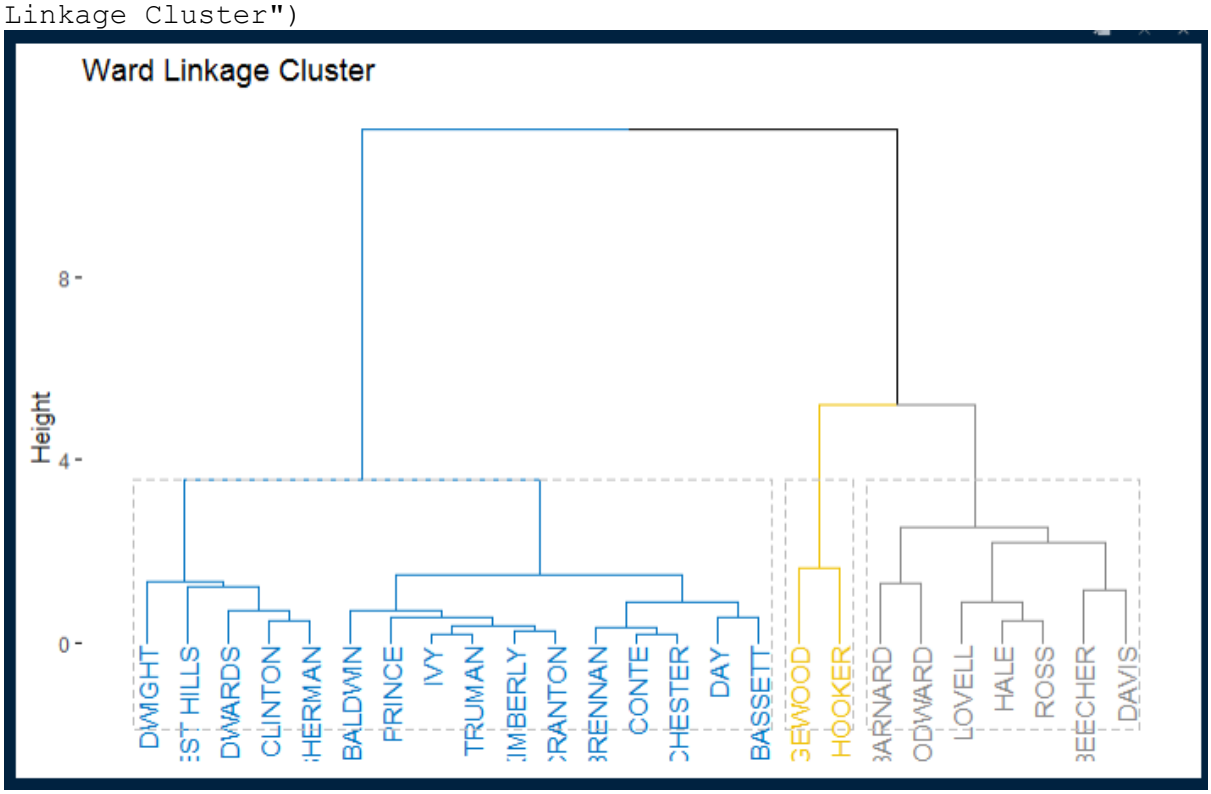

```
fviz_dend(Clus2_Ave, k = 2, k_colors = "jco", rect = T, main =
"Average Linkage Cluster")
```



```
fviz_dend(Clus2_Cen, k = 3, k_colors = "jco", rect = T, main =
"Centroid Linkage Cluster")
```

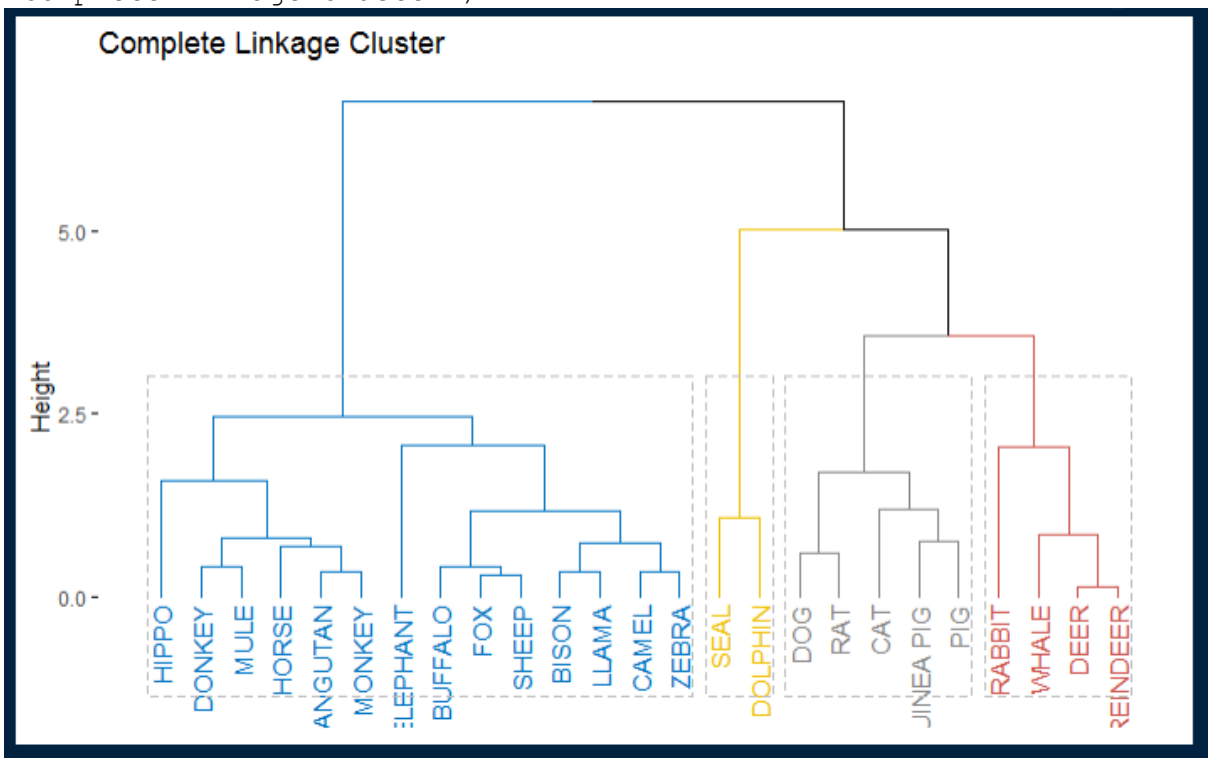



```
fviz_dend(Clus2_War, k = 3, k_colors = "jco", rect = T, main = "Ward
Linkage Cluster")
```

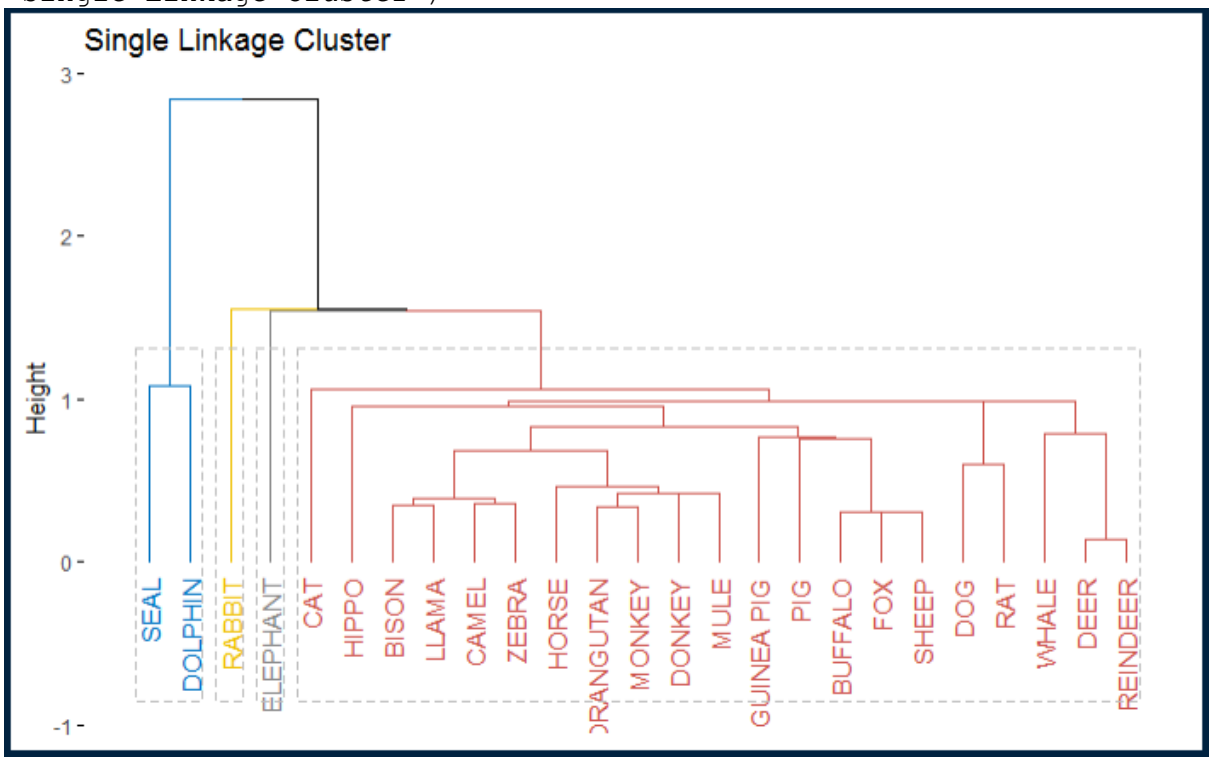



```
```

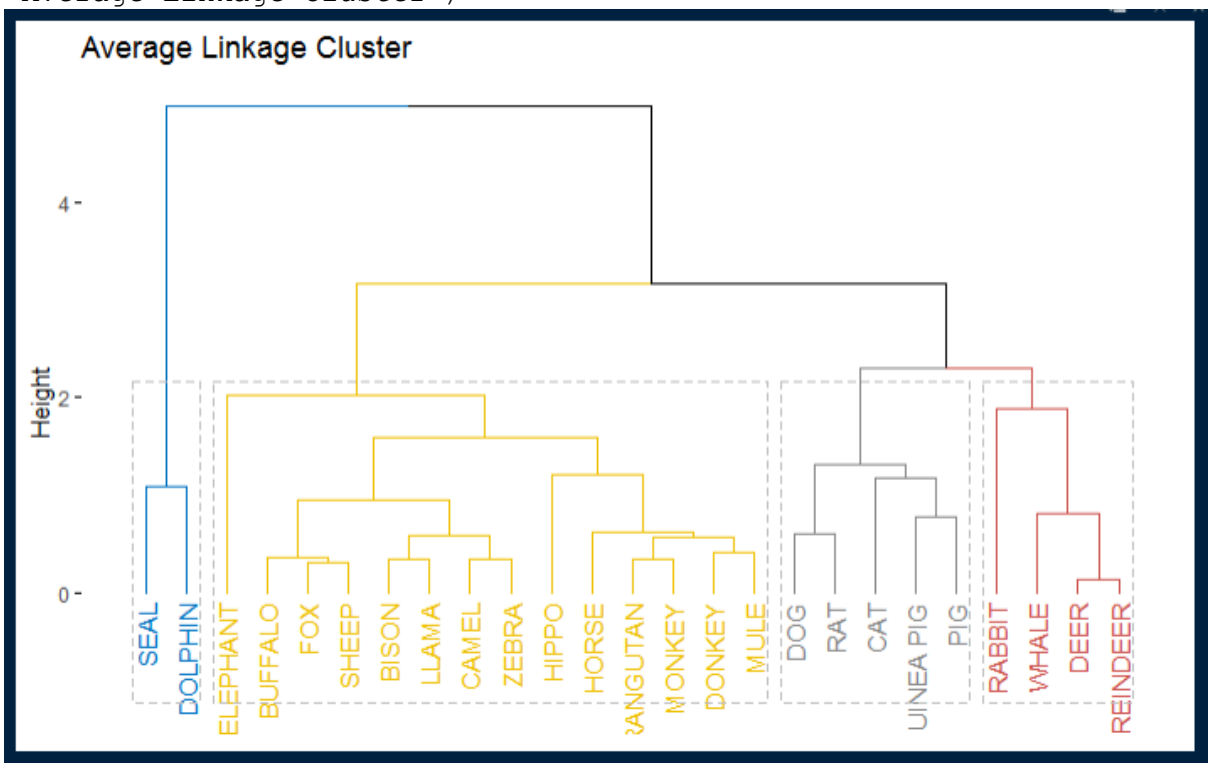
```{R}
#Dataset 3
fviz_dend(Clus3_Com, k = 4, k_colors = "jco", rect = T, main =
"Complete Linkage Cluster")
```



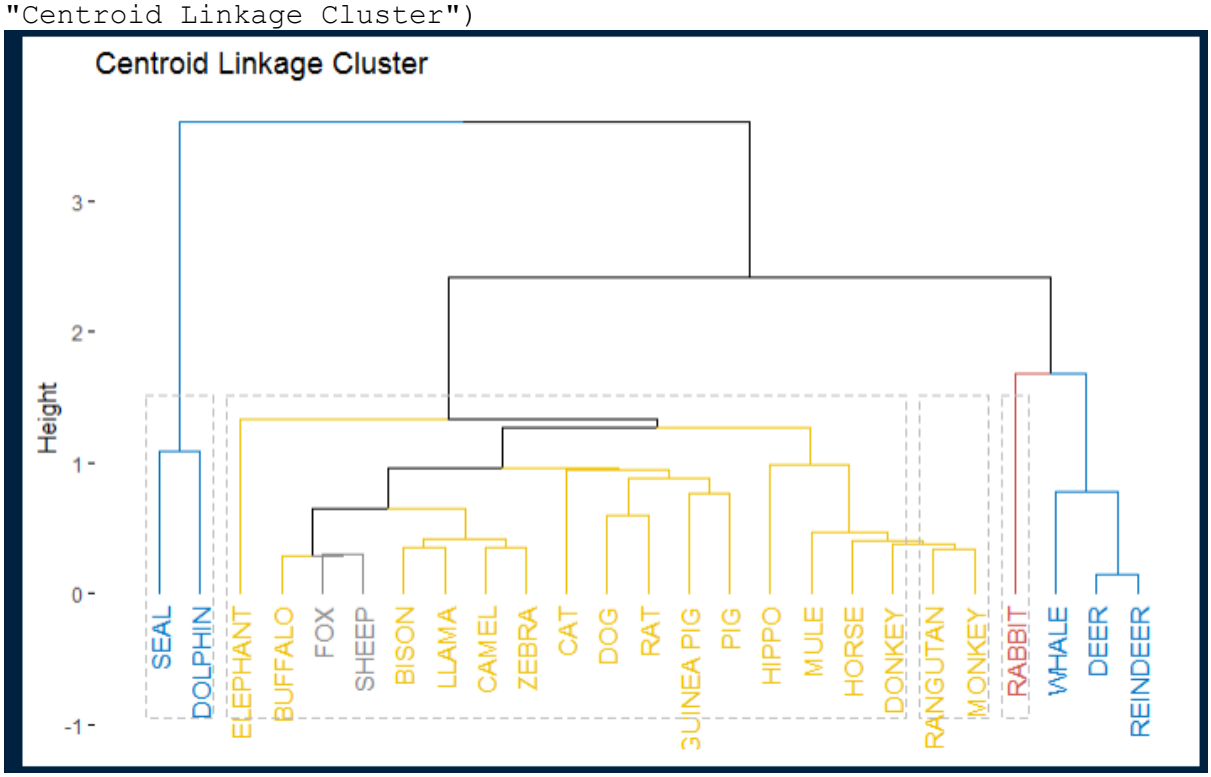

```
fviz_dend(Clus3_Sin, k = 4, k_colors = "jco", rect = T, main =
"Single Linkage Cluster")
```



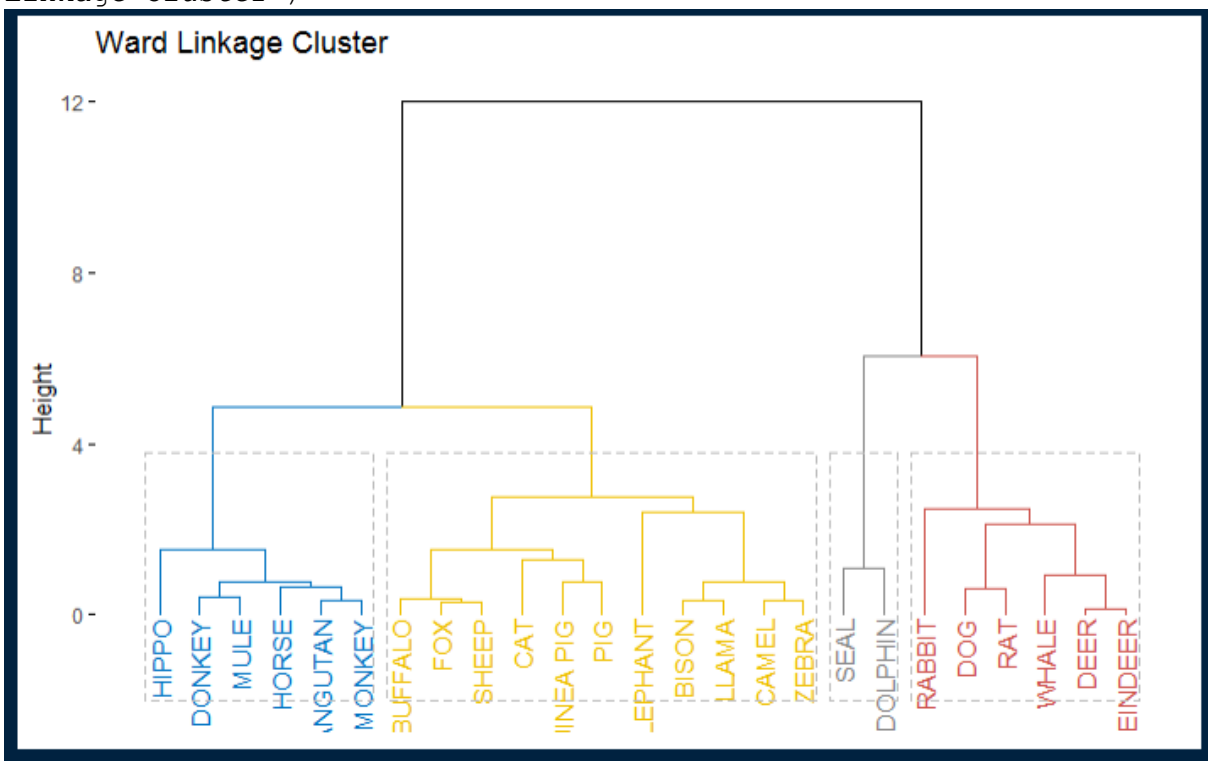
```
fviz_dend(Clus3_Ave, k = 4, k_colors = "jco", rect = T, main =
"Average Linkage Cluster")
```




```
fviz_dend(Clus3_Cen, k = 4, k_colors = "jco", rect = T, main =
"Centroid Linkage Cluster")
```



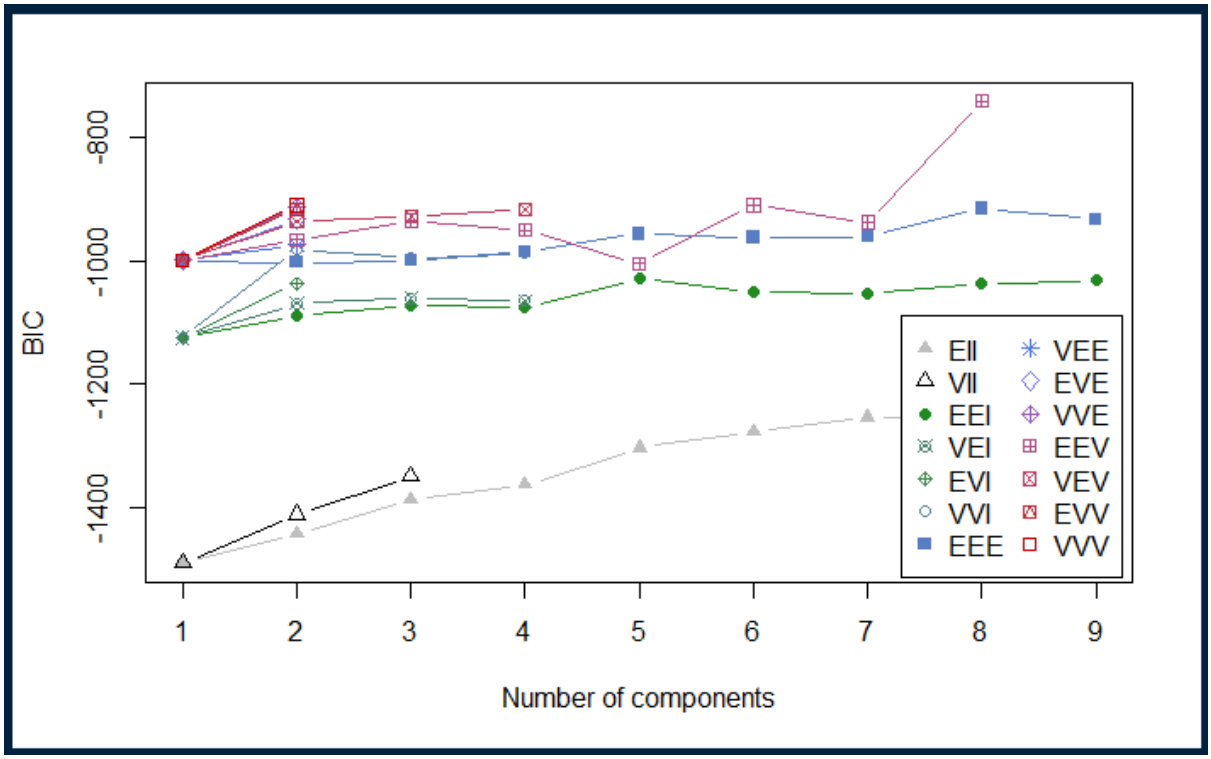
```
fviz_dend(Clus3_War, k = 4, k_colors = "jco", rect = T, main = "Ward Linkage Cluster")
```



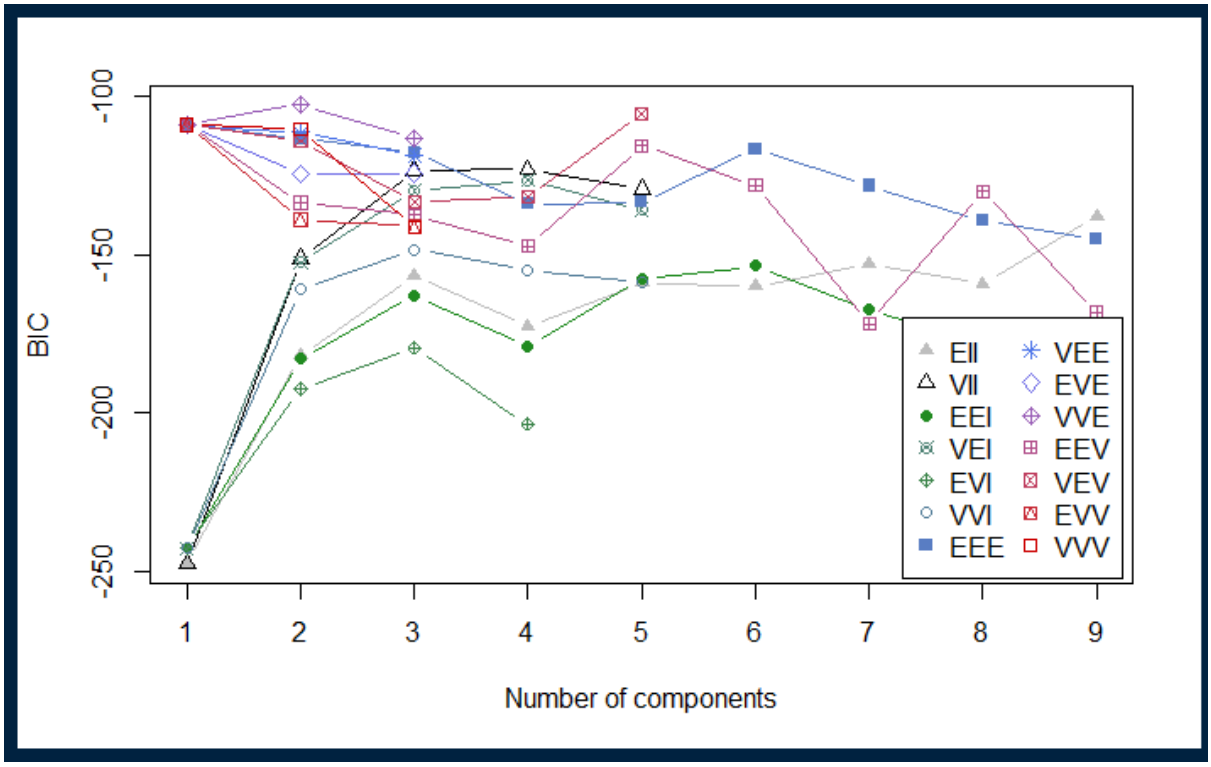

```
```

c. Tentukan nilai k (banyaknya klaster) berdasarkan nilai BIC atau yang lain

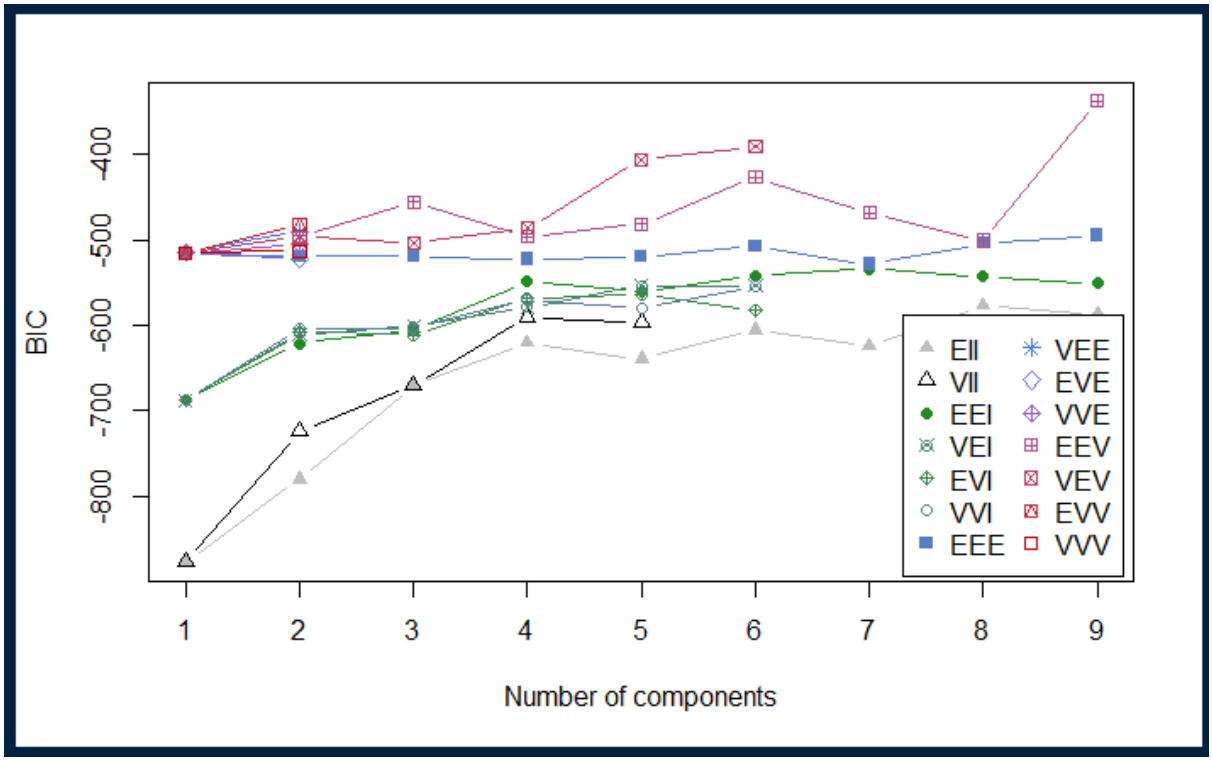
```{R}
#Dataset 1
library (mclust)
Clust1 <- as.matrix(nutrient)
Cluster1<-Mclust(nutrient)
summary(Cluster1)
plot(Cluster1)
```




```{R}
#Dataset 2
library (mclust)
Clust2 <- as.matrix(achieve)
Cluster2<-Mclust(achieve)
summary(Cluster2)
plot(Cluster2)
```




```{R}
#Dataset 3
library (mclust)
Clust3 <- as.matrix(milk)
Cluster3<-Mclust(milk)
summary(Cluster3)
plot(Cluster3)
```



d. Buatlah table rekapitulasi

Dataset 1 : nutrient

No	Metode	Nomer klaster	Anggota klaster
1	Complete-Linkage	1	7
		2	2
		3	1
		4	17
2	Single-Linkage	1	2
		2	1
		3	1
		4	23
3	Average-Linkage	1	2
		2	1
		3	1
		4	23
4	Centroid	1	4
		2	9
		3	1
		4	12
5	Ward	1	7
		2	2
		3	1
		4	17

Dataset 2 : achieve

No	Metode	Nomer klaster	Anggota klaster
1	Complete-Linkage	1	2
		2	15
		3	8
2	Single-Linkage	1	23
		2	1
		3	1
3	Average-Linkage	1	20
		2	5
4	Centroid	1	12
		2	1
		3	12
5	Ward	1	16
		2	2
		3	7

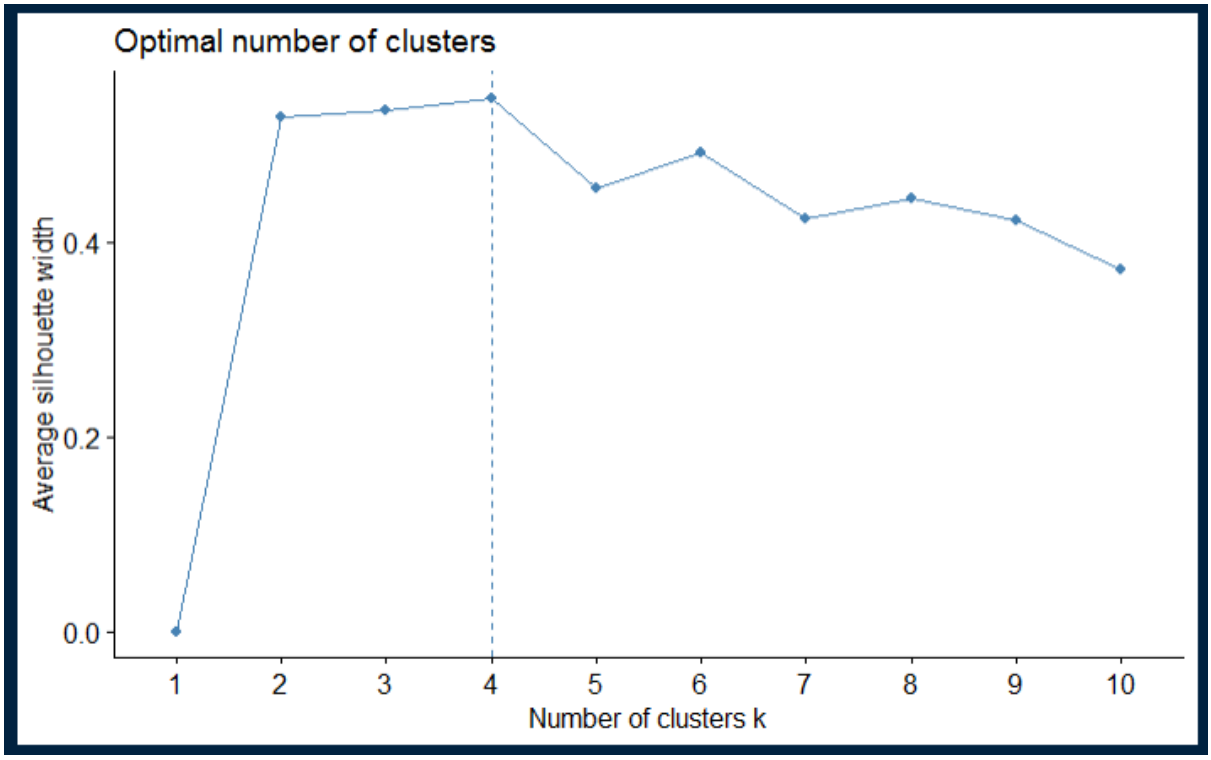
Dataset 3 : milk

No	Metode	Nomer klaster	Anggota klaster
1	Complete-Linkage	1	14
		2	2
		3	5
		4	4
2	Single-Linkage	1	2
		2	1
		3	1
		4	21
3	Average-Linkage	1	2
		2	14
		3	5
		4	4
4	Centroid	1	5
		2	18
		3	2
		4	1
5	Ward	1	6
		2	11
		3	2
		4	6

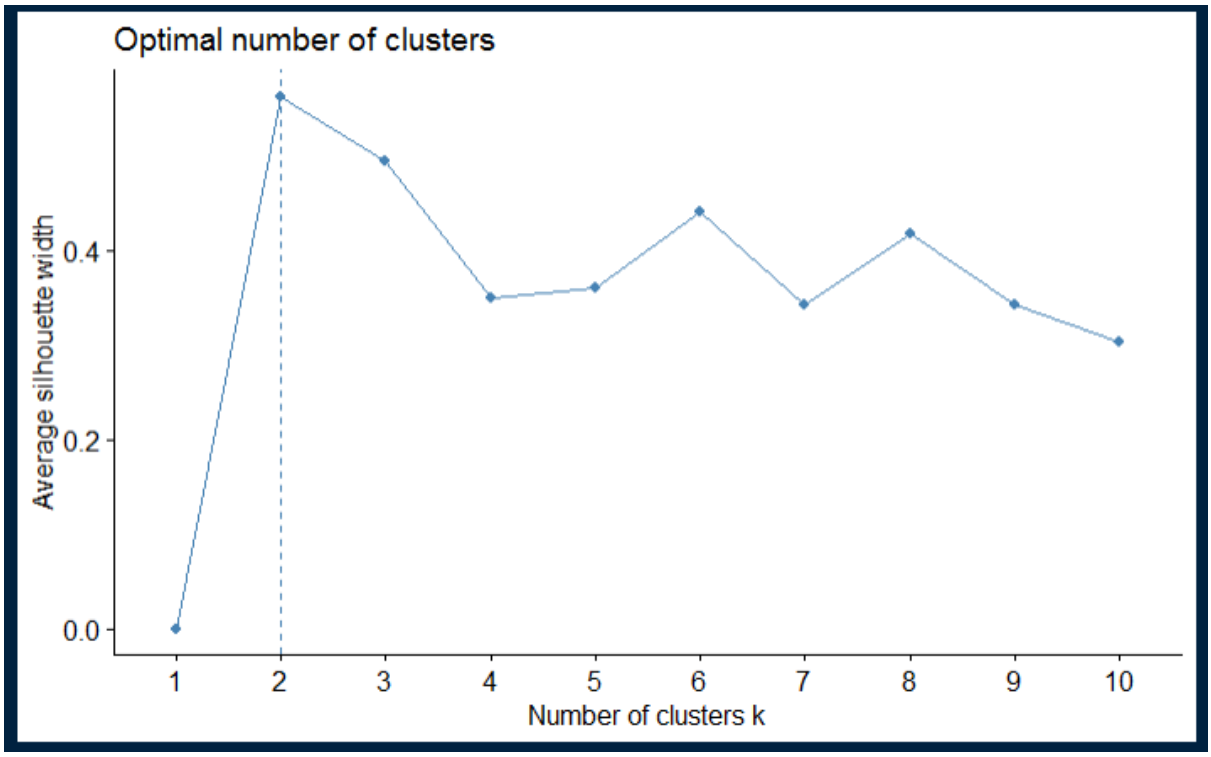
3. Gunakan no 1 untuk metode-metode berikut :
    a. K-means dan buatlah table rekapitulasi seperti no 1d.
    b. K-medoids dan buatlah table rekapitulasi seperti no 1d.
    c. K-medians dan buatlah table rekapitulasi seperti no 1d.

```{R}
#datasets 1
fviz_nbclust(nutrient, kmeans, method = "silhouette")
kmean1 <- kmeans(nutrient, 2)
kmean1

#datasets 2
fviz_nbclust(achieve, kmeans, method = "silhouette")
kmean2 <- kmeans(achieve, 2)
kmean2

#datasets 3
fviz_nbclust(milk, kmeans, method = "silhouette")
kmean3 <- kmeans(milk, 2)
kmean3
```

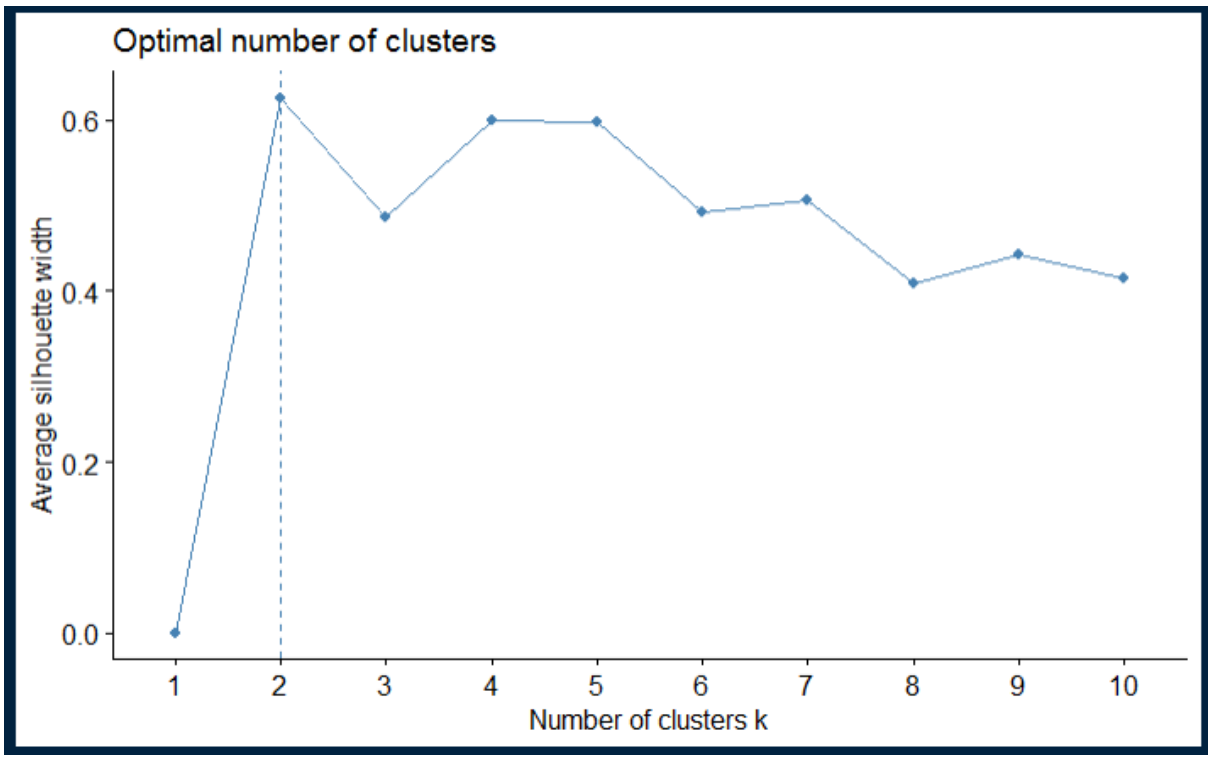
Datasets 1



Datasets 2



Datasets 3



Datasets	Metode	Nomer klaster	Anggota klaster
1	K-Means	1	13
		2	12
2	K-Means	1	12
		2	14
3	K-Means	1	11
		2	14